Docket Number: P-LG4878

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of PADGETT et al.	)	Examiner: Gary BENZION
	)	
	)	Group Art Unit: 1634
	)	
Serial No.: 10/066,390	)	I hereby certify that this correspondence
	)	is being deposited with the United States Postal service
	)	with sufficient postage for first class mail in a package
	)	addressed to Commissioner for Patents, P.O. Box 1450,
	)	Alexandria, VA 22313-1450
Filed: February 2, 2002	)	on October 26, 2007
	)	
For: A Method of Increasing Complementarity in a Heteroduplex	)	by Wayne Fitzmaurice
	)	Wayne P. Fitzmaurice
	)	Reg. No. 58,274

DECLARATION UNDER RULE 132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

October 26, 2007

Dear Sir:

I, Hal S. Padgett, do hereby declare and state:

1. I received a B.S. in Life Sciences in1988 from the University of Missouri at Rolla and a Ph.D. in Molecular Microbiology and Microbial Pathogenesis in 1996 from Washington University in Saint Louis, Missouri. I joined Large Scale Biology Corporation in 1998. For the past nine years my responsibilities involved viral vector development and molecular evolution programs at Large Scale Biology Corporation.

2. I am a named inventor of the subject matter that is claimed and for which a patent is sought on the invention entitled A Method of Increasing Complementarity in a

Heteroduplex.

3. The methods described in this application yield products that are surprisingly and unexpectedly different from that which is obtained simply from treatment of heteroduplexes with a DNA repair system.

4. The following data serves to further illustrate the results obtained in experiments of the type described in Examples 3, 4, 6, and 10 in the specification.

5. The series of experiments described in Examples 3 and 4 of the specification were performed to demonstrate that the Genetic Reassortment by Mismatch Resolution (GRAMMR) method efficiently creates shuffled gene sequences from linear heteroduplex molecules. Although the results of this experiment have already been presented in the specification, the results are shown in greater detail in Figure A as a series of graphical DNA alignments of the shuffled progeny genes with the parent genes to further illustrate the nature of the output from the claimed process and to contrast those results with results obtained from exposing heteroduplex DNA to a DNA repair system.

In this experiment, linear heteroduplex molecules were created by annealing two single-stranded DNAs of opposite strandedness generated by PCR from the two different parent genes described in Example 3 of the application. A sample of this heteroduplex DNA preparation was incubated in the presence of the GRAMMR reaction components, as described in Example 3, prior to cloning and sequencing. The negative control for the experiment consisted of a sample in which the GRAMMR reaction components were omitted, but was otherwise treated in parallel to the GRAMMR-treated heteroduplex sample. This negative control was included to measure the background level of sequence recombination that occurs when the non-GRAMMR-treated heteroduplex is exposed to a DNA repair system upon introduction into *E. coli* cells.

Figure A shows a graphical representation of the DNA sequences of randomly chosen output molecules from the experiment. The group of sequences from the negative control

reaction is depicted in the top panel of Figure A and the group of sequences derived from the CEL I treated samples are shown in the bottom panel of Figure A. In each figure each sequence is depicted in two ways: a 'nucleotide view' and a 'reassortment view' as shown. In the 'nucleotide view', gray represents nucleotides that are in common between the two parent molecules and dark blue or green represents nucleotides that are specific to one or the other parent. Light blue and red represent mutations that may occasionally arise, usually as a result of PCR amplification. As one follows the representation of a particular sequence across the panel, one can see the sequence switching between dark blue and green representing information transfer events and corresponding sequence reassortment. The results are easier to visualize in bulk with the representation of the same sequences shown at the bottom portion of each figure panel. In this 'reassortment view' the red dots show the midpoint between areas of the sequence that have switched from one parent sequence to the other (blue to green) or vice versa. Each of these is analogous to a recombinational crossover event.

The overall results of this shuffling experiment are described in Example 4 of the application and are summarized here. Only two of 10 clones derived from the negative control showed sequence recombination, with each of those having only a single crossover event (Figure A, top panel). Relatively infrequent crossover events among the negative controls are typical in our hands and are presumed to be caused either by the effects of the *E. coli* DNA repair system as described previously (1, 2, 3, 4, 9, 10) or, because the DNAs were PCR amplified prior to cloning, by 'jumping PCR' as described by Paabo (7).

In contrast to the negative controls, 100% of the GRAMMR-treated samples were shuffled (Figure A, bottom panel). Additionally, these samples displayed an average of 5 crossovers per clone which is equivalent to roughly nine per kilobase of heteroduplex region in the substrate. Both the percentage of shuffled clones and the amount of sequence reassortment are far higher in the GRAMMR-treated samples than in the DNA repair system-treated samples (negative controls).

6. Another experiment is described to illustrate the use of the GRAMMR method and to highlight the differences between results from the GRAMMR process and results from the use of a DNA repair system to treat heteroduplexes.

This experiment was previously described in Example 6 of the application. Heteroduplex DNAs used as substrates in the experiment were generated by restriction enzyme digestion of the parental plasmid pBSWTGFP (SEQ ID NO: 03) with KpnI and the parental plasmid pBSC3GFP (SEQ ID NO: 04) with NgoM-IV followed by spin column purification of the linearized plasmids, mixing and heat dissociation of the complementary strands and annealing of the resulting single stranded DNAs to one another to form duplexed DNAs. Because of the staggered cut sites in the original plasmids, the annealed DNAs forming heteroduplexes between strands derived from the different parent DNAs can assume a circular topology whereas duplexes resulting from re-annealing of perfectly matched parent DNA strands (from a single parent) remain linear. The fact that circular plasmid DNA molecules can transform *E. coli* with markedly higher efficiency than linear molecules provides a positive bias toward transformation and recovery of molecular clones that are derived originally from the circular heteroduplex molecules that are the desired target of this process.

DNA samples containing heteroduplex DNAs prepared as above were used as substrate in Genetic Reassortment by Mismatch Resolution (GRAMMR) reactions in which a replicate series of 10 microliter reactions was prepared with each containing 5 microliters of the heteroduplex substrate, 1X NEB ligase buffer, 0.5 mM each dNTP, 2.0 units *E. coli* DNA ligase (NEB), and 1.0 unit of T4 DNA polymerase, in the presence of various concentrations of CEL I. The CEL I was from a cloned preparation and the amount used varied from 0.11 microliters at the highest, through a series of four, 3-fold serial dilutions. The only difference among individual reactions in this experiment was the amount of CEL I added.

The negative control for this experiment consisted of a final replicate in the series that received buffer instead CEL I, but was otherwise identical to the above series of

experimental samples. This control reaction was performed to monitor the background level of sequence reassortment that occurs at a low frequency when heteroduplex substrates are introduced into *E. coli* and are presumably directly acted upon by the DNA repair system of the host cell.

After one hour of incubation of all reactions at 25 degrees centigrade, one microliter aliquots of each were used to transform competent DH5-alpha *E. coli* that were then plated on solid medium plates. A number of resulting colonies from the agar plates corresponding to each reaction were randomly picked and inoculated to liquid cultures. Plasmid DNA was extracted from each of the cultures followed by DNA sequence analysis of the complete GFP gene in each clone. The nucleotide sequences of these clones were analyzed by comparison to the two original parent GFP genes encoded by pBSWTGFP and pBSC3GFP. The results of this analysis are displayed in Figure B.

Figure B shows a graphic representation of the DNA sequences of output molecules from the experiment. The group of sequences from the negative control reaction is depicted in the top panel of Figure B and the group of sequences derived from the CEL I treated samples are shown in the bottom panel of Figure B. In each panel each sequence is depicted in two ways: a 'nucleotide view' and a 'reassortment view' as shown. In the 'nucleotide view', gray represents nucleotides that are in common between the two parent molecules and dark blue or green represents nucleotides that are specific to one or the other parent. Light blue and red represent mutations or sequencing artifacts that are occasionally observed. As one follows the representation of a particular sequence across the panel, one can see the sequence switching between dark blue and green representing information transfer events and corresponding sequence reassortment. The results are easier to visualize in bulk with the representation of the same sequences shown at the bottom of each figure panel. In this 'reassortment view' the red dots show the midpoint between areas of the sequence that have switched from one parent sequence to the other (blue to green) or vice versa. Each of these is analogous to a recombinational crossover event.

Figure B, top panel depicts the results from the negative control showing only one recombined clone in ten. This is a typical result in which the few recombinants that are observed in the negative controls show only a single or double crossover. Such results presumably reflect the level of sequence reassortment that is caused by the action of the cell's DNA repair system on the heteroduplex substrate and are consistent to what has been reported previously (1, 2, 3, 4, 9, 10).

In contrast to the negative control samples, molecular clones derived from heteroduplex DNAs exposed to complete GRAMMR reactions that included the CEL I mismatch endonuclease enzyme were extensively shuffled by only a single round of the GRAMMR process. As is shown in Figure B, bottom panel, the number of progeny clones bearing shuffled sequences is much higher in the CEL I treated samples than in the negative controls containing all reaction components but with CEL I omitted (eight of ten samples vs. one of ten, respectively).

From these data, it is evident that the GRAMMR process yields a reaction product that differs from what is obtained by simple exposure of a heteroduplex DNA to a DNA repair system.

7. The following experiment was described in Example 10 of the application and was performed to demonstrate that the claimed process is capable of reassortment of divergent gene sequences. In addition, the experiment provides further support for the ability of the method to create shuffled gene libraries from treatment of linear heteroduplexes as claimed in the present application. This experiment also highlights the contrast between the results obtained from the GRAMMR method and those following exposure of the heteroduplex to a DNA repair system.

The sequences being shuffled are homologous genes that share about 75% nucleotide identity. A linear heteroduplex DNA preparation was made and either treated using the GRAMMR process or used as a negative control by omitting the CEL I mismatch

endonuclease enzyme from the reaction as described in Example 10. The results of this experiment are shown in Figure C.

Figure C shows a graphic representation of the DNA sequences of randomly chosen output molecules from the experiment. The group of sequences from the negative control reaction is depicted in the top panel of Figure C and the group of sequences derived from the CEL I treated samples are shown in the bottom panel of Figure C. In each panel each sequence is depicted in two ways: a 'nucleotide view' and a 'reassortment view' as shown. In the 'nucleotide view', gray represents nucleotides that are in common between the two parent molecules and dark blue or green represents nucleotides that are specific to one or the other parent. Light blue and red represent mutations that may occasionally arise, presumably as a result of PCR amplification. As one follows the representation of a particular sequence across the panel, one can see the sequence switching between dark blue and green reflecting the occurrence of information transfer events and corresponding sequence reassortment. The results are easier to visualize in bulk with the representation of the same sequences shown at the bottom of each figure panel. In this 'reassortment view' the red dots show the midpoint between areas of the sequence that have switched from one parent sequence to the other (blue to green) or vice versa. Each of these is analogous to a recombinational crossover event.

The overall results of this experiment are described in Example 10 of the application and are summarized here. None of the seven negative control clones shown in the top panel of Figure C showed any sequence recombination.

In contrast, seven out of eight of the GRAMMR-treated samples had been shuffled (Figure C, bottom panel) with an average of about 2.5 crossovers per clone.

The results of this experiment demonstrate the contrast between the effects of treating heteroduplex DNA substrates with the claimed invention versus the effects of treatment of the heteroduplex with a DNA repair system.

Additional experiments are described here that are not included in the examples of the application since they had not been performed at that time. However, these examples provide additional support for the difference between the claimed invention and the use of a DNA repair system to create shuffled DNA molecules. The negative controls are treated the same in each case, and are therefore useful for assessing the level of sequence reassortment caused by treatment with a DNA repair system.

8. Another experiment is described to further illustrate the use of the GRAMMR method and to highlight the differences between results from the GRAMMR process and results from the use of a DNA repair system to treat heteroduplexes.

This experiment was performed by the method described in Example 6 of the application and section 6 of this declaration. Heteroduplex DNAs used as substrates in the experiment were generated by restriction enzyme linearizing the parental plasmid pGW-U1MP with Stu I and the parental plasmid pGW-ToMVMP with Sma I followed by spin column purification of the linearized plasmids, mixing and heat dissociation of the complementary strands and annealing of the resulting single stranded DNAs to one another to form duplexed DNAs. Because these parental plasmids differed in the approximately 800 base pair movement protein (MP) gene regions, but not the rest of the plasmid, the mismatch-containing region of DNA was limited to the MP gene. The MP genes share about 75% nucleotide identity.

The DNA samples containing heteroduplex DNAs were used as substrate in Genetic Reassortment by Mismatch Resolution (GRAMMR) reactions in which a replicate series of 10 microliter reactions was prepared with each containing 1 microliter of the heteroduplex substrate, 1X NEB ligase buffer, 50 mM KCl, 0.5 mM each dNTP, 2.0 units *E. coli* DNA ligase (NEB), and 1.0 unit of T4 DNA polymerase, in the presence of various concentrations of CEL I. The CEL I was from a cloned preparation and the amount used varied from 0.33 microliters at the highest, through a series of four, 3-fold serial dilutions. The only difference among individual reactions in this experiment was the amount of CEL I added.

The negative control for this experiment consisted of a final replicate in the series that received buffer instead CEL I, but was otherwise identical to the above series of experimental samples. This control reaction was performed to monitor the background level of sequence reassortment that has been reported to occur at a low frequency when heteroduplex substrates are introduced into *E. coli* and are directly acted upon by the DNA repair system of the host cell (1, 2, 3, 4, 9, 10).

After one hour of incubation of all reactions at 25 degrees centigrade and 30 minutes on ice, one microliter aliquots of each were used to transform competent DH5-alpha strain *E. coli* that were then plated on solid medium plates. A number of resulting colonies from the agar plates corresponding to each reaction were randomly picked and inoculated to liquid cultures. Plasmid DNA was extracted from each of the cultures followed by DNA sequence analysis of the complete movement protein gene in each clone. The nucleotide sequences of these clones were analyzed by comparison to the two original parent genes encoded by pGW-U1MP and pGW-ToMVMP. The results of this analysis are displayed in Figure D.

Figure D shows a graphic representation of the DNA sequences of output molecules from the experiment. The group of sequences from the negative control reaction is depicted in the top panel of Figure D and the group of sequences derived from the CEL I treated samples are shown in the bottom panel of Figure D. In each figure each sequence is depicted in two ways: a 'nucleotide view' and a 'reassortment view' as shown. In the 'nucleotide view', gray represents nucleotides that are in common between the two parent molecules and dark blue or yellow represents nucleotides that are specific to one or the other parent. Light blue represents mutations or sequencing artifacts that are occasionally observed. As one follows the representation of a particular sequence across the panel, one can see the sequence switching between dark blue and yellow representing information transfer events and corresponding sequence reassortment. The results are easier to visualize in bulk with the representation of the same sequences shown at the bottom of each figure panel. In this 'reassortment view' the red dots show the midpoint

between areas of the sequence that have switched from one parent sequence to the other (blue to yellow) or vice versa. Each of these is analogous to a recombinational crossover event.

The negative control sample in which the heteroduplexes were only exposed to the cellular repair system gave zero recombined clones out of 12 that were sequenced. Eight of these are depicted in Figure D, top panel. In sharp contrast, molecular clones derived from heteroduplex DNAs exposed to complete GRAMMR reactions that included the CEL I mismatch endonuclease enzyme were extensively shuffled in the single round of the process with six of eight clones being shuffled (Figure D, bottom panel). In addition, the frequency of sequence reassortment events caused by the GRAMMR process, with an average of more than eight in each clone, is very high.

From these data, it is evident that the GRAMMR process yields a reaction product that differs from what is obtained by simple exposure of a heteroduplex DNA to a DNA repair system.

9. Another experiment is described to further illustrate the use of the GRAMMR method and to highlight the differences between results from the GRAMMR process and results from the use of a DNA repair system to treat heteroduplexes.

This experiment was performed by the method described in Example 6 of the application and section 6 of this declaration. Heteroduplex DNAs used as substrates in the experiment were generated by restriction enzyme digestion of the parental plasmid pBSWTGFP with NgoM-IV and the parental plasmid pBSC3BFP, a blue fluorescent variant of pBSC3GFP, with KpnI followed by spin column purification of the linearized plasmids, mixing and heat dissociation of the complementary strands and annealing of the resulting single stranded DNAs to one another to form duplexed DNAs.

The DNA samples containing heteroduplex DNAs were used as substrate in Genetic Reassortment by Mismatch Resolution (GRAMMR) reactions in which a replicate series

of 10 microliter reactions was prepared with each containing 1 microliter of the heteroduplex substrate, 1X NEB ligase buffer, 0.5 mM each dNTP, 2.0 units *E. coli* DNA ligase (NEB), and 1.0 unit of T4 DNA polymerase, in the presence of various concentrations of CEL I. The CEL I was from a cloned preparation and the amount used varied from 0.11 microliters at the highest, through a series of four, 3-fold serial dilutions. The only difference among individual reactions in this experiment was the amount of CEL I added.

The negative control for this experiment consisted of a final replicate in the series that received buffer instead CEL I, but was otherwise identical to the above series of experimental samples. This control reaction was performed to monitor the background level of sequence reassortment that has been reported to occur at a low frequency when heteroduplex substrates are introduced into *E. coli* and are directly acted upon by the DNA repair system of the host cell (1, 2, 3, 4, 9, 10).

After one hour of incubation of all reactions at 25 degrees centigrade, one microliter aliquots of each were used to transform competent XL1-Blue strain *E. coli* that were then plated on solid medium plates. A number of resulting colonies from the agar plates corresponding to each reaction were randomly picked and inoculated to liquid cultures. Plasmid DNA was extracted from each of the cultures followed by DNA sequence analysis of the complete GFP gene in each clone. The nucleotide sequences of these clones were analyzed by comparison to the two original parent genes encoded by pBSWTGFP and pBSC3BFP. The results of this analysis are displayed in Figure E.

Figure E shows a graphic representation of the DNA sequences of output molecules from the experiment. The group of sequences from the negative control reaction is depicted in the top panel of Figure E and the group of sequences derived from the CEL I treated samples are shown in the bottom panel of Figure E. In each figure each sequence is depicted in two ways: a 'nucleotide view' and a 'reassortment view' as shown. In the 'nucleotide view', gray represents nucleotides that are in common between the two parent molecules and dark blue or green represents nucleotides that are specific to one or the

other parent. Light blue represents mutations or sequencing artifacts that are occasionally observed. As one follows the representation of a particular sequence across the panel, one can see the sequence switching between dark blue and green representing information transfer events and corresponding sequence reassortment. The results are easier to visualize in bulk with the representation of the same sequences shown at the bottom of each figure panel. In this 'reassortment view' the red dots show the midpoint between areas of the sequence that have switched from one parent sequence to the other (blue to green) or vice versa. Each of these is analogous to a recombinational crossover event.

The negative control sample in which the heteroduplexes were only exposed to the cellular repair system gave zero recombined clones out of 14 that were sequenced. Ten of these are depicted in Figure E, top panel. In sharp contrast, molecular clones derived from heteroduplex DNAs exposed to complete GRAMMR reactions that included the CEL I mismatch endonuclease enzyme were extensively shuffled in the single round of the process with nine of 10 clones being shuffled (Figure E, bottom panel). In addition, the frequency of sequence reassortment events caused by the GRAMMR process, with an average of more than nine in each clone, is very high.

From these data, it is evident that the GRAMMR process yields a reaction product that is dramatically different from what is obtained by simple exposure of a heteroduplex DNA to a DNA repair system.

10. From these data, it is my opinion that the GRAMMR process yields shuffled molecular products that are different from those obtained from simply exposing heteroduplexes to a DNA repair system. The output polynucleotides from the negative controls in the accompanying figures result from exposure of heteroduplexes to a DNA repair system, as has already been amply described in the literature (1, 2, 3, 4, 9, 10). These controls therefore serve as a valid measure of the effects of a DNA repair system on the heteroduplex. These controls consistently show only low levels of sequence reassortment (an occasional recombined molecule, typically with only one crossover

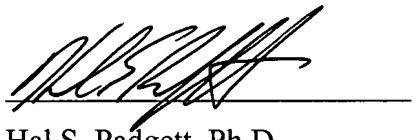
point) which is consistent with what has been reported previously when heteroduplexes with multiple mismatches were exposed to DNA repair systems either *in vivo* (1, 2, 3, 4, 9, 10) or *in vitro* (5, 6). DNA repair systems have been observed to perform similarly *in vitro* to the way they do *in vivo* (5, 8).

In marked contrast with results of the negative controls, the current invention produces very high frequencies of shuffled clones, usually with a high density of sequence reassortment events. If the GRAMMR shuffling method we describe were simply equivalent to a DNA repair system, it would be expected that similar results would be obtained from the claimed invention and from the negative controls. However, the results we get from the two are dramatically different, highlighting the contrast between the method of our invention and treatment with a DNA repair system.

These results are relevant to Claims 66-72, 78-83, 85, and 87-105 of the application.

11. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information are believed to be true, and that the statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, and patent issuing thereon, and patent to which this verified statement is directed.

October 26, 2007
Date



Hal S. Padgett, Ph.D.

- **Attachments**: Reference Listing
  Copies of References
  Figures A-E

References:

1. Abastado, J.P., et al., *Processing of complex heteroduplexes in Escherichia coli and Cos-1 monkey cells.* Proc Natl Acad Sci U S A, 1984. **81**(18): p. 5792-6.

2. Cami, B., P. Chambon, and P. Kourilsky, *Correction of complex heteroduplexes made of mouse H-2 gene sequences in Escherichia coli K-12.* Proc Natl Acad Sci U S A, 1984. **81**(2): p. 503-7.

3. Chang, S., et al., *Recombination following transformation of Escherichia coli by heteroduplex plasmid DNA molecules.* Gene, 1984. **29**(3): p. 255-61.

4. Fishel, R. and R. Kolodner, *Gene conversion in Escherichia coli: the recF pathway for resolution of heteroduplex DNA.* J Bacteriol, 1989. **171**(6): p. 3046-52.

5. Fishel, R.A. and R. Kolodner, *An Escherichia coli cell-free system that catalyzes the repair of symmetrically methylated heteroduplex DNA.* Cold Spring Harb Symp Quant Biol, 1984. **49**: p. 603-9.

6. Glazer, P.M., et al., *DNA mismatch repair detected in human cell extracts.* Mol Cell Biol, 1987. **7**(1): p. 218-24.

7. Paabo, S., D.M. Irwin, and A.C. Wilson, *DNA damage promotes jumping between templates during enzymatic amplification.* J Biol Chem, 1990. **265**(8): p. 4718-21.

8. Su, S.S., et al., *Mispair specificity of methyl-directed DNA mismatch correction in vitro.* J Biol Chem, 1988. **263**(14): p. 6829-35.

9. Sung, W.L., et al., *Simultaneous synthesis of human-, mouse- and chimeric epidermal growth factor genes via 'hybrid gene synthesis' approach.* Nucleic Acids Res, 1986. **14**(15): p. 6159-68.

10. Weber, H. and C. Weissmann, *Formation of genes coding for hybrid proteins by recombination between related, cloned genes in E. coli.* Nucleic Acids Res, 1983. **11**(16): p. 5661-9.

Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 5792–5796, September 1984
Genetics

Processing of complex heteroduplexes in *Escherichia coli* and *Cos*-1 monkey cells

(repair/gene conversion/diversity/histocompatibility gene sequences)

JEAN-PIERRE ABASTADO*, BRIGITTE CAMI*, TAM HUYNH DINH†, JEAN IGOLEN†, AND PHILIPPE KOURILSKY*

*Unité de Biologie Moléculaire du Gène, Equipe de Recherche No. 201 du Centre National de la Recherche Scientifique et Service Commun No. 20 de l'Institut National de la Santé et de la Recherche Médicale, and †Unité de Chimie Organique, Equipe de Recherche Associée No. 927 du Centre National de la Recherche Scientifique, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cédex 15, France

Contributed by Marianne Grunberg-Manago, May 7, 1984

ABSTRACT **Upon transformation into *Escherichia coli* or *Cos*-1 monkey cells, heteroduplex DNA made of two sequences containing many nucleotide mismatches yields a wide array of different molecules, some with a patchwork structure. Thus, complex heteroduplexes can be processed to generate many genetic variants.**

Genomic DNA, particularly in the higher eukaryotes, contains redundant sequences that share significant, but only partial, homology. It has often been suggested that such partially homologous sequences (e.g., genes of some multigene families) can undergo reciprocal and nonreciprocal genetic exchanges (recombination and conversion) (see refs. 1–3 for review; refs. 4–9). Such exchanges are likely to involve the formation of hybrid DNA (reviewed in refs. 10–13), which, as we emphasized earlier (1, 14), should contain many nucleotide mismatches. Similar structures could also arise as the result of mistakes in the replication process. It is thus important to learn about the *in vivo* behavior of such complex heteroduplexes.

Recently, we constructed heteroduplexes by annealing complementary strands of two different cloned mouse *H-2* genes in which about 8% of the base pairs were mismatched. When transfected into *Escherichia coli*, these heteroduplexes were frequently processed (repaired), resulting in *H-2* sequences composed of fragments of either parental gene (14). We now extend this observation to show that processing of complex heteroduplexes also takes place in *Cos*-1 monkey cells. We have studied a large number of processed molecules produced both in *E. coli* and *Cos*-1 monkey cells and observed a broad variety of molecules carrying patchworks of the parental sequences. Thus, as previously suspected by us (1, 14) and others (refs. 2, 15–17; M. Gefter and M. Fox, quoted in ref. 15), processing of complex heteroduplexes is a means of generating genetic diversity.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. Our standard *rec*$^+$ host (803 *sup*E *sup*F rk$^-$mk$^-$) and its *rec*A$^-$ (*rec*A13) derivative have been described (18). The other strains used are: C600 *rec*BC$^-$rk$^-$mk$^-$ (19), ES895, which is *str*A *trp*A540 *lac*Z(ICR36) *rec*A$^-$ *mut*S, and ES872, which is *str*A *trp*A540 *lac*Z(IRC36) *rec*A$^-$ *mut*L (20). Plasmids pH-2^d-1 and pH-2^d-3 and their sequences have been described (14). Their derivatives harboring the simian virus 40 (SV40) origin of replication were constructed by replacing the *Nde* I–*Ava* I 872-base-pair (bp) fragment from pBR322 by the *Nde* I–*Pvu* II 690-bp fragment from SV40.

Preparation and Transformation of Heteroduplexes. Experiments in *E. coli* were carried out as described (14). For the transformation of *Cos*-1 cells, heteroduplexes were prepared with different sets of restriction enzymes (see text) and transfected by the $CaPO_4$ procedure (21) into 2×10^7 *Cos*-1 monkey cells (22). Hirt supernatant (23) was then digested by *Dpn* I and transformed into *rec*A$^-$ *E. coli*.

Oligonucleotide Probes and *in Situ* Hybridization. Oligonucleotides were synthetized by the phosphotriester method in solid phase (24) from dimers and trimers, with triisopropylsulfonylnitrotriazole as the coupling agent (25).

The six oligonucleotides (numbered 1–6) are: (1) 5′ C-A-T-C-A-C-C-C-C-A-G-A-*T*-C-T; (2) 5′ T-G-C-C-*A*-T-G-T-G-*G*-A-*A*-C-A-T; (3) 5′ *C*-C-G-T-T-C-A-C-*T*-*G*-A-*C*-*T*-C-T; (4) 5′ T-*G*-G-*A*-C-T-G-*T*-A-C-T-G-*A*-G-T; (5) 5′ T-G-A-G-*T*-*C*-T-G-*C*-G-G-G-*C*-T-C; (6) 5′ T-G-A-*G*-C-T-*G*-C-A-*G*-C-T-C-C-T. Italicized letters show mismatched residues when annealed to the alternative *H-2* sequence. Oligonucleotides 1, 3, 4, and 6 are specific for pH-2^d-3, whereas oligonucleotides 2 and 5 are specific for pH-2^d-1, as experimentally demonstrated in control hybridization experiments (see below).

In situ hybridization was done as described elsewhere (26, 27).

RESULTS

Experimental Design. In this and our previous analysis (14), we have selected for study two plasmids, pH-2^d-1 and pH-2^d-3, which carry two cDNA inserts, 1.15 and 1.0 kilobases long, cloned in the *Pst* I site of pBR322 in the same orientation (14). The cDNA inserts correspond to transcripts of the mouse *H-2* multigene family that encodes polymorphic class I transplantation antigens (28). Their coding capabilities are irrelevant here. The aligned sequences (Fig. 1) of the two cDNAs differ by 86 nucleotides (14); in addition, pH-2^d-3 carries a 3-bp insertion with regard to pH-2^d-1, whereas pH-2^d-1 carries a 9-bp insertion and extends 142 bp further at the 5′ end. To prepare the heteroduplexes, one plasmid was cut with a pair of enzymes (*Eco*RI and *Hin*dIII) and the other one was cut with another pair (*Bam*HI and *Sph* I). Both sets cut the tetracycline resistance gene (Tc^R) of pBR322 in a nonoverlapping way, such that, upon denaturation and reannealing, heteroduplexes alone can generate an active Tc^R gene (14).

Such heteroduplex DNA was transformed into *E. coli* or *Cos*-1 monkey cells. In the latter case, DNA was extracted after 2 days and plasmids were recovered by cloning in *E. coli* (see below). Thus, in both approaches, Tc^R *E. coli* transformants were selected at the end. Reisolated clones were screened by *in situ* hybridization with six oligonucleotide probes specific for one or the other cDNA insert under proper hybridization conditions (26) and located as shown in Fig. 1. A typical experiment is shown in Fig. 2. On occasion, some hybridization signals were weak and difficult to inter-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); SV40, simian virus 40; Tc^r, tetracycline resistance.

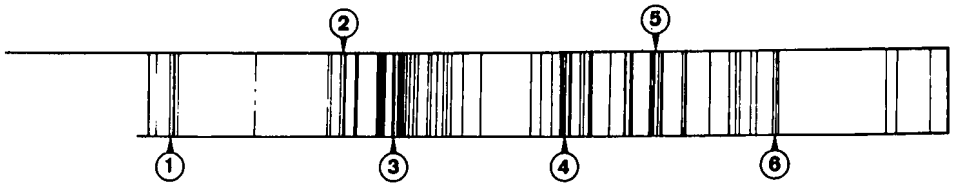


FIG. 1. Comparison of the nucleotide sequences of pH-2^d-1 and pH-2^d-3 cDNA inserts. Both sequences have been published (7). The two cDNA inserts are aligned, pH-2^d-1 on the top and pH-2^d-3 on the bottom, with the 5' end to the left. Mismatches are shown as vertical bars. The localization of the six oligonucleotide probes (numbered 1–6) is indicated.

pret and confirmation was sought by repeating the hybridization or by restriction mapping (or both). Due to lack of specific restriction sites, all structures could not be verified by mapping. When doubt persisted, the clone was discarded. In addition, we found that clones hybridizing with all probes harbored both plasmids as a recombinant dimer, whereas clones hybridizing with none of the oligonucleotide probes carried deleted plasmids. These major artifacts and the dubious clones are excluded from the results presented below.

Heteroduplex Processing in *E. coli* and Several *E. coli* Mutants. Heteroduplex DNA was transformed into several hosts: our standard *rec*$^+$ *E. coli*, its *rec*A$^-$ derivative, another recombination-deficient host with a *rec*BC$^-$ mutation, and two *rec*A$^-$ hosts carrying an additional *mut*L or *mut*S mutation known to interfere with the correction of single base-pair mismatches formed during *E. coli* DNA replication (29). Two hundred to 300 Tc^R clones were analyzed in each case. The majority had a hybridization pattern identical to that of pH-2^d-1 or pH-2^d-3 and were classified as "parental." In all *E. coli* hosts examined, 5–25% of the Tc^R clones displayed a pattern distinct from both parents and were tabulated as "rearranged." Hybridization results for 197 rearranged molecules are given in Fig. 3, where it is seen that they fall into a broad variety of hybridization patterns.

Fate of Heteroduplexes Introduced into *Cos*-1 Monkey Cells. To determine whether similar events could occur in mammalian cells, the same heteroduplexes were introduced into monkey cells in another vector that allows replication in monkey cells. The amplification of plasmid DNA in monkey cells was required to obtain sufficient quantities to transform *E. coli* and perform the molecular analysis of *H-2* sequences. Therefore, we removed from pH-2^d-1 and pH-2^d-3 a piece of DNA containing at least part of the so-called "poison sequence" that interferes with replication in monkey cells (30). In its place, we substituted a 690-bp fragment of the SV40 genome that contains the viral origin of replication (see *Materials and Methods*). The resulting plasmids, pH-2^d-1-SV and pH-2^d-3-SV, have acquired the ability to replicate in *Cos*-1 monkey cells, which express SV40 T antigen constitutively (22). Heteroduplexes were formed by using two sets of restriction enzymes (*Eco*RI and *Cla* I; *Bam*HI and *Sal* I) as above. In addition, it was necessary to ensure that any rearrangement observed after cloning back into *E. coli* had occurred in the mammalian cells and not in bacteria. Therefore, DNA extracted from *Cos*-1 cells 2 days after transformation was treated with the enzyme *Dpn* I, which cuts methylated G-A-T-C sites. The methyl groups are present in DNA grown in *E. coli* but are lost upon replication in mammalian cells (31). pH-2^d-1-SV, pH-2^d-3-SV, and heteroduplexes carry 21, 20, and 18 *Dpn* I sites, respectively. A *Dpn* I treatment should thus destroy all unreplicated heteroduplex DNA from the initial DNA preparation that might be present in, or contaminate, the mammalian extract (the effectiveness of this treatment was checked with an internal control). Finally, we transformed a *rec*A$^-$ *E. coli* host to avoid recombination between two plasmids that might have entered the same bacterium.

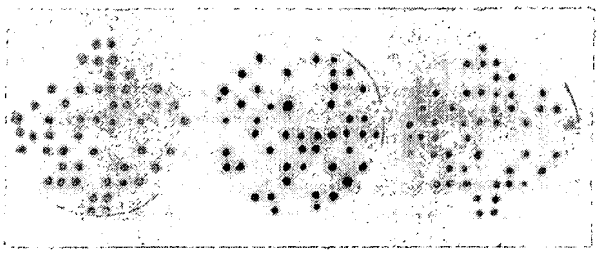


FIG. 2. *In situ* colony hybridization with oligonucleotide probes. One hundred Tc^R clones in the *rec*$^+$ host were hybridized *in situ* with oligonucleotides 1 (*Left*), 2 (*Center*), and 3 (*Right*).

Three-hundred twenty-seven Tc^R transformants were obtained and analyzed by *in situ* hybridization. About half (174) had a parental type, with a bias in favor of pH-2^d-3-SV (126 vs. 48 of the pH-2^d-1-SV type). Of the remaining clones, 39 were smaller and 46 were larger than the parental molecules, some being multimers, indicative of recombination or DNA insertion. Sixty-eight had the parental size, out of which 17 were discarded because their structure was unclear and could not be cross-checked.

The hybridization patterns of the 51 remaining clones are given in Fig. 3. Again, their distribution was scattered among the various possible types. Restriction maps that were constructed for 13 clones confirmed their patchwork structures (Fig. 4). Clone p237 was partially sequenced in the region of the strand switches and the nucleotide sequence confirmed the structure (Fig. 4 and data not shown). The entire sequence of p292 was determined and revealed additional complexity (Figs. 4 and 5). Although hybridization with the six probes had revealed four strand switches, the sequence demonstrated eight switches, one of which is associated with an insertion of 19 nucleotides replacing 6 nucleotides of the normal sequence (Fig. 5). These 19 nucleotides represent a duplication followed by an inversion of nucleotides of the parental sequences, the generation of which will be discussed elsewhere (unpublished data).

DISCUSSION

We had previously shown processing of complex heteroduplexes in *rec*A$^-$ *E. coli* (14). Seven processed molecules were identified by restriction mapping of plasmids isolated from random transformants. Large populations could not be screened by this procedure. *In situ* colony hybridization with a set of oligonucleotide probes is a more powerful approach, which we have exploited to: (*i*) further document correction in *E. coli*, (*ii*) extend some of our observations to heteroduplexes introduced into animal cells, and (*iii*) show that processing generates many different variants, thereby being a possible source of genetic diversity.

Heteroduplex Processing in *E. coli*. Our results with *rec*A$^-$ and *rec*BC$^-$ mutants demonstrate that at least part of the rearranged molecules are not produced by classical recombination.

Therefore, heteroduplex repair appears as the most likely processing mechanism. However, two mutants deficient in

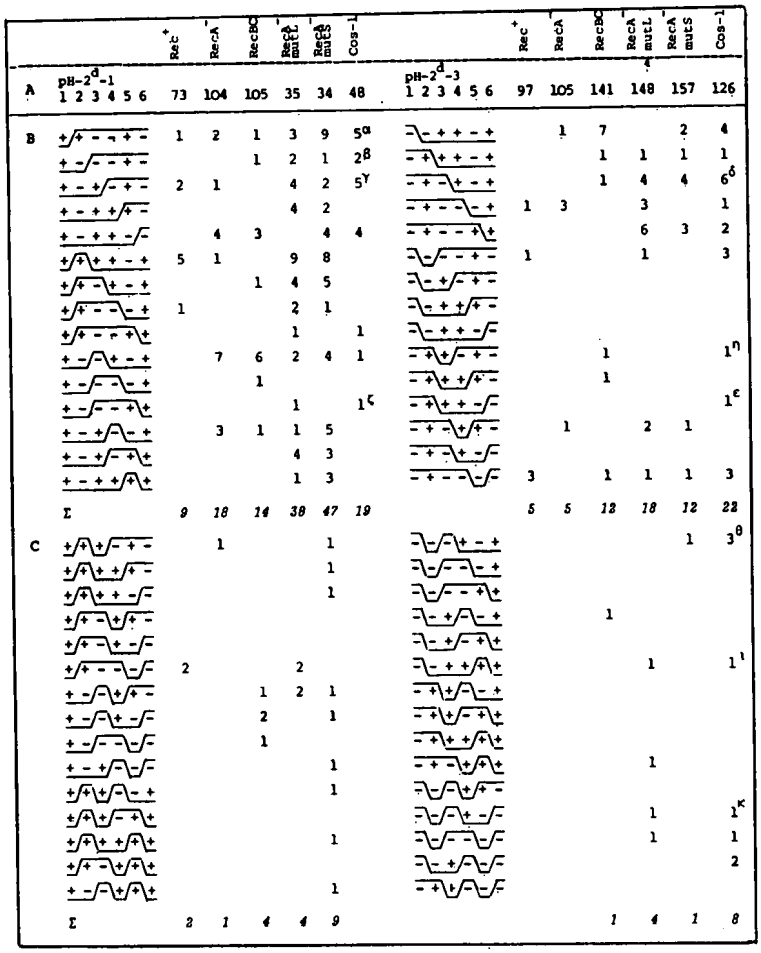


FIG. 3. Hybridization pattern of rearranged clones. In *E. coli*, the transformation efficiency, measured with pBR322 DNA, was in the order of 3–8 × 10^6 Tc^R clones per μg, the *rec*$^+$ host being systematically 1.5–2 times more efficient than the others. With heteroduplex DNA, there were 2.8 × 10^4, 1 × 10^4, 7.5 × 10^3, 3 × 10^3, and 4.5 × 10^3 Tc^R transformants per μg in *rec*$^+$, *recA*$^-$, *recBC*$^-$, *recA*$^-$*mut*L, and *recA*$^-$*mut*S, respectively. Rearranged clones represented 6%, 10%, 10.5%, 24%, and 22% of the clones examined (200–300 in each case), respectively. The experiment using *Cos*-1 cells is described in the text. The number of clones obtained that hybridize to oligonucleotide probes 1–6 is shown. (*A*) Parental types (pH-2^d-1 to the left, pH-2^d-3 to the right); (*B*) "simple" rearranged types implying one- or two-strand switches, which can be interpreted as the result of a single repair event (see text); (*C*) more complex rearranged types involving three or more strand switches. In *B* and *C*, sums in the various columns are given as Σ. The figure is organized symmetrically with all molecules having the 5′ part of the pH-2^d-1 and pH-2^d-3 sequences to the left and right, respectively. The restriction maps of some of these clones are represented in Fig. 5. Those clones are: α, p31 and p232; β, p255 and p309; γ, p89; δ, p23 and p201; ε, p237; ζ, p176; η, p160; θ, p12; ι, p187; and κ, p292.

the repair of single base-pair mismatches (*mut*L$^-$ and *mut*S$^-$) also display processing. But the mutations could be leaky (20) or the *mut*L and *mut*S products may be dispensable in the phenomenon examined, perhaps because of the large number of mismatched residues. In fact, the *recA*$^-$ *mut*L and *recA*$^-$ *mut*S strains yield a high proportion of rearranged molecules and could prove useful to produce genetic variants by processing of complex heteroduplexes. (Note that the proportion of rearranged molecules given in Fig. 3 is almost certainly underestimated, since the set of oligonucleotides could not detect all variations.) Relatively fewer rearranged molecules were produced in the *recA*$^+$ host. Since more Tc^R clones were usually obtained per μg of DNA (legend to Fig. 3), it is possible that a higher background of Tc^R transformants was generated by recombination (14). Alternatively, *rec*A protein could interfere with processing, perhaps through one of its many activities on single- or double-stranded DNA (or both) (reviewed in ref. 32).

Hybridization results in Fig. 3 have been presented symetrically with the simplest structures at the top (Fig. 3*B*) and the more complex ones at the bottom (Fig. 3*C*). It is seen that most of the rearranged molecules can be interpreted as resulting from a single repair event involving two strand switches (one possibly being in the plasmid sequence) (Fig. 3*B*). There also exist more complex patterns indicative of two (or more) repair tracts (Fig. 3*C*). In addition, there are more rearranged molecules with the 5′ sequence of pH-2^d-3 than with that of pH-2^d-1 (left vs. right of Fig. 3). This suggests that the large loop caused by the extra sequence in pH-2^d-1 is preferentially eliminated.

In many of the molecules, the rearrangement is revealed by at least two oligonucleotide probes, implying that a block of sequence, rather than a single nucleotide, is corrected. This agrees with our previous analyses (partially based on DNA sequencing) (14). The precise mechanism is unknown but could involve nick-translation. Our heteroduplex molecules are nicked in the Tc gene where some repair events may be initiated (unpublished data). Molecules with two or more strand switches within the cDNA sequence must, however, involve an internal rearrangement. A thorough analysis of the processing mechanism(s) with the present system may prove difficult because, contrary to some phage models (29, 33, 34), the role of DNA replication is difficult to assess.

Heteroduplex Processing in *Cos*-1 Monkey Cells. Similar experiments, with appropriate modifications (see *Results*) were carried out in *Cos*-1 monkey cells. Contrary to *E. coli*, no recombination-deficient cell mutants could be used. In preliminary experiments in which *Cos*-1 cells were transformed by both circular plasmids (pH-2^d-1-SV and pH-2^d-3-SV), we did observe a few recombinants in the cDNA inserts (data not shown). Some rearranged molecules may have been produced by recombination between parental molecules segregated after replication of the heteroduplex strands rather than by repair.

Nevertheless, generating the patchwork structure of plasmid p292 (Fig. 5) would require eight crossing over in 500 bp. This, and the high incidence of rearranged molecules, suggests that at least some of them are produced by heteroduplex repair. In fact, this is, perhaps, the first strong evidence of heteroduplex repair in mammalian cells.

Insertions and deletions are frequently observed. Some of these events may be irrelevant to the heteroduplex structure

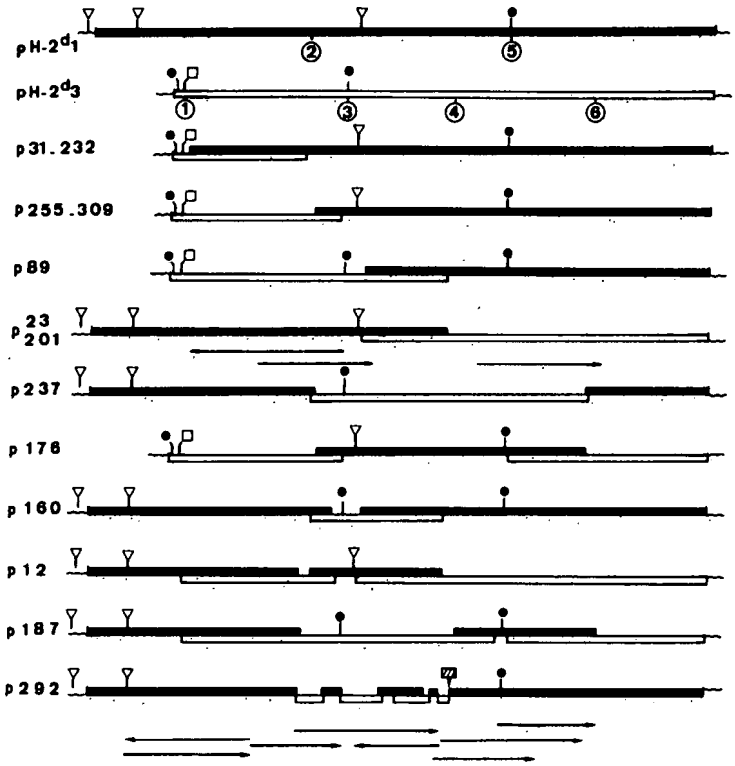


FIG. 4. Restriction maps of 13 plasmid molecules rearranged in *Cos*-1 cells. The pH-2^d-1 and pH-2^d-3 sequences are shown as filled and open bars, respectively. The six oligonucleotide probes map as indicated. Restriction maps were established with *Bgl* II (□), *Hin*fI (●), and *Hpa* II (▽). The structure of the clones deduced from the data is depicted as segments of filled and open bars representing the parental sequences. The bars overlap where data do not permit further discrimination. Arrows indicate the sequencing strategy and the sequenced areas of p237 and p292. The hatched box shows the 19-nucleotide inversion in p292.

because it has recently been shown that bacterial markers (in a homoduplex form) suffer a very high rate of punctual and nonpunctual mutations when shuttled into *Cos*-1 cells and back into *E. coli* (35, 36). These variations will, therefore, add to those actually resulting from heteroduplex correction. A high repair activity in transformed cells would not be surprising, since it could be induced by the introduction of large amounts of DNA or the transformation procedure *per se* (or both) (37).

Generation of Diversity. Fig. 3 shows a broad diversity of processed molecules. Patterns of processing may not be completely random: there seem to be "hot spots" in certain *E. coli* hosts and, perhaps, in *Cos*-1 cells (Fig. 3). They will not be discussed here because, in our opinion, the very complexity of the starting material precludes any reliable correlation between these patterns and specific features of the heteroduplex (loops, potential methylation sites in *E. coli*, etc.). (It should be noted, however, that the DNA used for transfection of monkey cells was grown in *E. coli* and may have distinct modifications.)

A priori, mismatched nucleotides could be corrected as adjacent blocks, or one independently of the other. Our results, both in *E. coli* and *Cos*-1 cells, clearly show block correction. This is in line with a body of previous genetic evidence in several biological systems (e.g., *Saccharomyces cerevisiae* and *Ascobolus*) in which co-correction of adjacent markers was frequently observed (38, 39). If repair is the basic mechanism of processing, the length of repaired tracts in our experiments is variable, but often smaller than previously reported for simpler heteroduplexes (40–43, 16). Thus, repair tracts may become shorter as heterology increases.

Molecules that have undergone a single processing (repair) event are tabulated in Fig. 3*B*. Of the 30 possible combinations of hybridization signals yielding this category of molecules, 25 have been observed in *E. coli* and 17 in *Cos*-1 cells, where the survey was more limited. We conclude that the borders of the corrected tracts are, at this level of resolution, located at random on the heteroduplex molecule. More generally, the number of possible variants (N) generated by a single block correction between two sequences with n mismatches is: $N = n(n + 1)/2$ (in the present case, $n = 86$ and $N = 3741$). Thus, such a mechanism is capable of generating considerable diversity (1, 16). From our observations, it also appears that several processing events affect the same molecule. This occasionally results in the formation of complex patchworks (such as that in p292), which adds to the diversity.

We have not shown that such complex heteroduplexes ac-

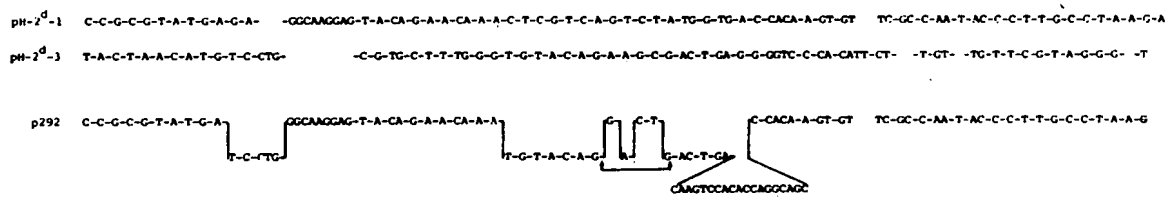


FIG. 5. Nucleotide sequence of p292. The nucleotide sequence of the cDNA insert of clone p292 is aligned with that of pH-2^d-1 and pH-2^d-3. Only those positions where pH-2^d-1 and pH-2^d-3 diverge are indicated. The correspondent positions in p292 are depicted on two lines depending on whether they fit pH-2^d-1 or pH-2^d-3. At all other positions (where pH-2^d-1 and pH-2^d-3 share the same nucleotide), p292 is identical, except for 19 nucleotides, C-A-A-G-T-C-C-A-C-A-C-C-A-G-G-C-A-G-C, which are complementary to the region of pH-2^d-1 and pH-2^d-3 indicated by arrows.

tually form *in vivo*. However, it is very likely that hybrid DNA is generated in the course of genetic exchanges based on homology (crossing-over and conversion) when the homology is sufficient (1, 32). Exploring the degree of heterology that allows heteroduplex formation *in vivo* is thus a major goal of our future analyses.

In summary, we have observed that complex heteroduplexes introduced into *E. coli* and *Cos*-1 monkey cells are processed with significant frequency. In *E. coli*, the use of recombination-deficient mutants strongly suggests that repair, rather than recombination, is involved, although the actual processing mechanism is not understood. In monkey cells, processing could be due to repair or recombination (or both). The finding that complex patchworks are occasionally formed suggests that repair is involved in at least some of the processing events. In both *E. coli* and monkey cells, we observe that the corrected areas are essentially random. We suspect that processing of complex heteroduplexes may happen in many living cells. It would then be a general mutational mechanism capable of modifying one, several, or many nucleotides at once, in a manner that will usually conserve the common traits in the primary sequence of the parental molecules (1). The rapid generation of diversity over evolutionary times may be particularly significant in eukaryotes, in which partially homologous sequences are abundant (1). More specifically, this mechanism may help in understanding genetic exchanges reported or postulated in many multigene families (2–9, 44).

We are very grateful to Dr. M. J. Masse for occasional help and useful criticisms, to Dr. P. Meulien for reviewing the manuscript, to Dr. P. Masiakowski (Strasbourg) for providing advice on the use of Whatman 540 paper in hybridization, and to many colleagues in and outside the laboratory, particularly Dr. M. Radman, for stimulating discussions. We are indebted to Mrs. V. Caput for her invaluable help in the preparation of the manuscript. This work was supported by grants of Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale (including Institut National de la Santé et de la Recherche Médicale Contract 124011 to J.I.), and the Fondation pour la Recherche Médicale Française.

1. Kourilsky, P. (1983) *Biochimie* **65**, 85–93.
2. Baltimore, D. (1981) *Cell* **24**, 592–594.
3. Petes, T. & Fink, G. R. (1982) *Nature (London)* **300**, 216–217.
4. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21**, 627–638.
5. Ollo, R. & Rougeon, F. (1983) *Cell* **32**, 515–523.
6. Pays, E., Van Assel, S., Laurent, M., Darville, M., Vervoort, T., Van Meirvenne, N. & Steinert, M. (1983) *Cell* **34**, 371–381.
7. Lalanne, J. L., Brégégère, F., Delarbre, C., Abastado, J. P., Gachelin, G. & Kourilsky, P. (1982) *Nucleic Acids Res.* **10**, 1039–1049.
8. Weiss, E. H., Mellor, A., Golden, L., Fahmer, K., Simpson, E., Hurst, J. & Flavell, R. A. (1983) *Nature (London)* **301**, 671–674.
9. Pease, L. R., Schulze, D. H., Pfaffenbach, G. M. & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 242–246.
10. Holliday, R. (1964) *Genet. Res.* **5**, 282–304.
11. Meselson, M. S. & Radding, C. M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 358–361.
12. Radding, C. M. (1978) *Annu. Rev. Biochem.* **47**, 847–880.
13. Fogel, S., Mortimer, R. K. & Lusnak, K. (1981) in *The Molecular Biology of the Yeast Saccharomyces*, eds. Strathern, J. N., Jones, E. W. & Broach, J. R. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 289–339.
14. Cami, B., Chambon, P. & Kourilsky, P. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 503–507.
15. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. (1978) *Science* **202**, 11–17.
16. Bourguignon-Van Horen, M. E., Bortcorn, A., Caillet-Fauquet, P., Diver, W. P., Dohet, C., Doubleday, O. P., Lecomte, P., Maenhaut-Michel, G. & Radman, M. (1982) *Biochimie* **61**, 559–564.
17. Brégégère, F. (1983) *Biochimie* **65**, 229–237.
18. Murray, N. Z., Brammar, W. J. & Murray, K. (1977) *Mol. Gen. Genet.* **150**, 53–61.
19. Rougeon, F., Kourilsky, P. & Mach, B. (1975) *Nucleic Acids Res.* **2**, 2365–2378.
20. Vaccaro, K. K. & Siegel, E. C. (1977) *Mutat. Res.* **42**, 443–446.
21. Wigler, M., Pellicer, A., Silverstein, S., Axel, R., Urlaub, G. & Chasin, L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1373–1376.
22. Gluzman, Y. (1981) *Cell* **23**, 175–182.
23. Hirt, B. (1967) *J. Mol. Biol.* **26**, 365–369.
24. Crea, R., Hirose, T. & Itakura, K. (1979) *Tetrahedron Lett.* **20**, 395–398.
25. de Rooij, J. F. M., Wille-Hazleger, G., Van Deursen, P. H., Serdijn, J., & van Boom, H. H. (1979) *Recl. Trav. Chim.* **98**, 537–548.
26. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6613–6617.
27. Reyes, A. A., Johnson, M. J., Schöld, M., Ito, H., Ike, Y., Morin, C., Itakura, K. & Wallace, R. B. (1981) *Immunogenetics (N.Y.)* **14**, 383–392.
28. Hood, L., Steinmetz, M. & Malissen, B. (1983) *Annu. Rev. Immunol.* **1**, 529–568.
29. Glickman, B. W. & Radman, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1063–1067.
30. Lusky, M. & Botchan, M. (1981) *Nature (London)* **293**, 79–81.
31. Peden, K. W. C., Pipas, J. M., Pearson-White, S. & Nathans, D. (1980) *Science* **209**, 1392–1395.
32. Radding, C. M. (1982) *Annu. Rev. Genet.* **16**, 405–437.
33. Lichten, M. & Fox, M. S. (1983) *Genetics* **103**, 5–22.
34. Pukkila, P. J., Peterson, J., Herman, G., Modrich, P. & Meselson, M. (1983) *Genetics* **104**, 571–582.
35. Razzaque, A., Mizusawa, H. & Seidman, M. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3010–3014.
36. Calos, M. P., Lebkowski, J. S. & Botchan, M. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3015–3019.
37. Colbère-Garapin, F., Garapin, A. C. & Kourilsky, P. (1982) in *Current Topics in Microbiology and Immunology*, eds. Hofschneider, P. H. & Goebel, W. (Springer, Berlin), Vol. 96, pp. 145–157.
38. Roman, H. & Fabre, F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6912–6916.
39. Nicolas, A. & Rossignol, J. L. (1983) *EMBO J.* **2**, 2265–2270.
40. Wildenberg, J. & Meselson, M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2202–2206.
41. Wagner, R., Jr., & Meselson, M. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 4135–4139.
42. Lai, C. J. & Nathans, D. (1975) *Virology* **66**, 70–81.
43. Miller, L. K., Cooke, B. E. & Fried, M. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3073–3077.
44. Stoeckert, C. Y., Jr., Collins, F. S. & Weissman, S. M. (1984) *Nucleic Acids Res.* **12**, 4469–4480.

Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 503–507, January 1984
Genetics

Correction of complex heteroduplexes made of mouse *H-2* gene sequences in *Escherichia coli* K-12

(repair/gene conversion/genetic diversity/H-2 antigens)

BRIGITTE CAMI, PHILIPPE CHAMBON, AND PHILIPPE KOURILSKY

Unité de Biologie Moléculaire du Gène, Equipe de Recherche no. 201 du Centre National de la Recherche Scientifique et Service Commun no. 20 de l'Institut National de la Santé et de la Recherche Médicale, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cédex 15, France

Communicated by Sydney Brenner, September 19, 1983

ABSTRACT **We have prepared heteroduplexes between two plasmids that carry, in the same orientation, two H-2 cDNA inserts, 1.15 and 1.0 kilobase long, respectively. Their sequences encode two distinct class I transplantation antigens of the mouse and differ by 8% of their nucleotides. Molecules with a rearranged array of restriction sites were found after transformation and cloning in an *Escherichia coli recA*$^-$ host. Nucleotide sequences showed that the rearranged molecules derived their nucleotides from the two parental strands. Thus, correction of these complex heteroduplexes takes place in *E. coli* and probably involves repair mechanisms. It provides the basis for a mutational process in which several nucleotides (amino acids) can be altered in a single event. It also offers a practical means of making genetic variants. Several other implications are discussed.**

Heteroduplexes can form *in vivo* by DNA strand exchange between partially homologous, but not identical, sequences (reviewed in ref. 1). They can also result from replication mistakes. In *Escherichia coli*, the newly synthesized strand, being transiently undermethylated, is preferentially corrected (reviewed in refs. 2 and 3). *E. coli dam*$^-$ mutants, deficient in a major methylation activity, display high mutation rates, as expected from random correction of either strand (4).

Heteroduplexes can be prepared *in vitro*, transformed into living cells, and their *in vivo* correction can then be studied. Such analyses have been carried out mostly in *E. coli* (5–8; see ref. 1 for review). With heteroduplexes of λ phage DNA carrying up to four nucleotide mismatches, Wagner and Meselson (9) observed independent correction as well as cocorrection of the marker mutations. Heteroduplexes of simian virus 40 (10, 11) and polyoma (12) mutants have been transfected into mammalian cells, where mismatch repair is also believed to take place.

These studies have been carried out with "simple" heteroduplexes, carrying one or a few nucleotide mismatches. Little is known about the correction of more complex structures, involving many noncomplementary nucleotides, which we, and others, suspect to be a mechanism capable of generating considerable diversity (3, 13, 14; see below). Because this idea may explain some of the genetic polymorphism in eukaryotic multigene families (14), particularly in the mouse *H-2* genes studied in our laboratory, we have undertaken an analysis of the fate of complex heteroduplexes. As a first step, these studies have been carried out in *E. coli*.

In the *H-2* multigene family, which encodes the polymorphic class I transplantation antigens, proteins and genes analyzed so far display high homology, with 80–95% of identical residues between any two aligned sequences (see refs. 15–18 for review). We have selected for study two blocks of *H-2* sequences, about 1 kilobase (kb) long, which differ in many positions, prepared heteroduplexes *in vitro*, and transformed them into *E. coli*. We report here that correction takes place, and we discuss several implications.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIAL AND METHODS

Bacterial Strains. The *recA*$^+$ and *recA*$^-$ *E. coli* strains used here were 803 *supE supF* $r^{k-}m^{k-}$ and 803 *supE supF* $r^{k-}m^{k-}$ *recA*$^-$ (19). The *recA*$^-$ strain has been periodically tested in this laboratory for UV sensitivity, formation of small colonies, and inability to support the growth of certain λ mutants. The *dam*$^-$ strain was gM82 *dam*$^-$ (4). Strains harboring pH-2^d-1 and pH-2^d-3 have been described (20, 21).

Enzymes and Isotopes. Restriction enzymes were purchased from New England BioLabs and Bethesda Research Laboratories and were used in the conditions recommended by the manufacturers. Polynucleotide kinase was from Boehringer Mannheim and terminal deoxynucleotidyl transferase from P-L Biochemicals. [γ-^{32}P]ATP and α-^{32}P-labeled cordycepin (specific activity, 3,000 Ci/mmol; 1 Ci = 37 GBq) were purchased from Amersham.

Formation and Transformation of Heteroduplexes. One plasmid (several micrograms) was digested by *Eco*RI and *Hin*dIII; the other was cut by *Bam*HI and *Sph* I. The plasmids were then extracted once by chloroform/isoamyl alcohol, precipitated by ethanol, and resuspended in 10 mM Tris·HCl, pH 7.5/1 mM EDTA, at a concentration of 250 μg/ml; 500 ng of each plasmid was mixed in a final volume of 10 μl of the same buffer. The mixture was denatured by boiling for 3 min in water. Annealing was for 4 hr at 63°C (22). The sample was then diluted 1:10 in 0.1 M Tris·HCl (pH 7.1) and aliquots containing 10–50 ng of DNA were transformed into *E. coli* (23).

DNA Sequence Analysis. Nucleotide sequences were determined as described by Maxam and Gilbert (24) using DNA fragments labeled by [γ-^{32}P]ATP and polynucleotide kinase, or ^{32}P-labeled cordycepin and terminal deoxynucleotidyl transferase (25).

RESULTS

Choice of Sequences. In the mouse, most somatic cells display at their surface three types of class I molecules, coded by distinct loci (*H-2D*, *K*, and *L*) of chromosome 17 (reviewed in refs. 15–18). We have isolated previously two cDNA clones, pH-2^d-1 and pH-2^d-3, that probably encode the *H-2D* and *L* products, respectively, in the *d* haplotype (20, 21). These cDNAs were cloned in the bacterial plasmid pBR322. The inserts are 1.15 and 1 kb long and represent incomplete copies of the 1,800-nucleotide long H-2 mRNAs, starting from poly(A) in 3'. They all encompass the third extracellular domain, the membrane spanning region, and the cytoplasmic COOH terminus of H-2 heavy chain, as well

Abbreviations: kb, kilobase(s); bp, base pair(s).

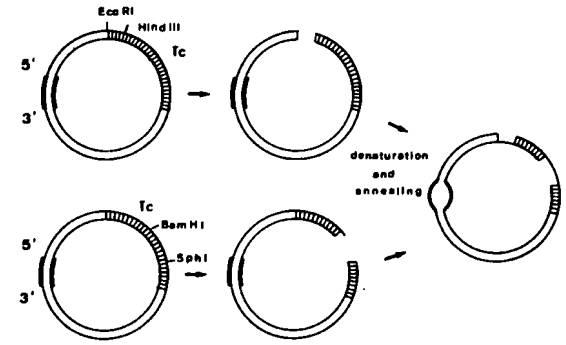


FIG. 1. Preparation of heteroduplexes. The figure depicts the formation of heteroduplexes with gaps in the *Tc* gene. Other molecules are formed in the annealing reaction, particularly homoduplexes of truncated plasmids and concatenates of heteroduplex molecules. Examination of the annealed mixture by electrophoresis indicated that, in our experimental conditions, the latter were much less abundant (≈20% or less) than circular heteroduplexes.

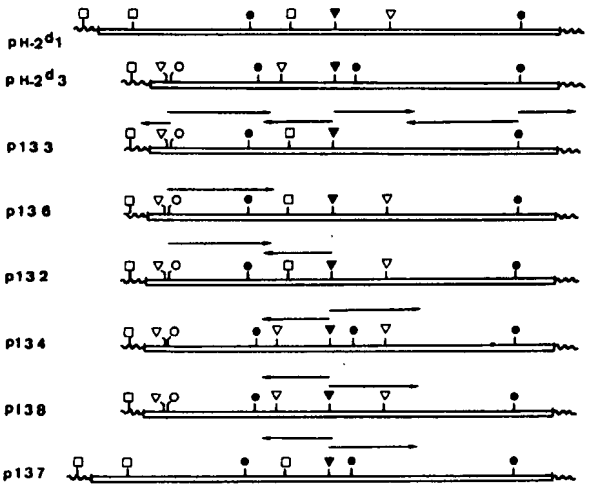


FIG. 2. Restriction maps of rearranged clones. Plasmid DNA, digested by one or several restriction enzymes, was subjected to electrophoresis in a 1% agarose gel. The inserts are shown as bars and plasmid DNA as a wavy line. The unique *Sac* I site present in all plasmids is shown as ▼. Other enzymes used are indicated as follows: *Bgl* II (○), *Hin*fI (▽), *Hpa* II (□), *Rsa* I (●). The regions in which sequences were determined are indicated by arrows, the origins of which correspond to the restriction sites used for terminal labeling.

as 480 base pairs (bp) of noncoding sequence downstream from the stop codon. We have selected for study pH-2^d-1 and pH-2^d-3 in which the inserts have the same orientation with regard to pBR322. Their nucleotide sequences can readily be aligned. They differ in 86 positions including a 3-bp deletion in pH-2^d-1 and a 9-bp deletion in pH-2^d-3. In addition, the pH-2^d-1 insert extends 142 bp further at the 5′ end. It also carries a longer poly(A) tract in the 3′ end (40 residues versus 30 in pH-2^d-3). The lengths of the G·C homopolymeric tails are roughly similar but have not been precisely determined.

Preparation and Transformation of Heteroduplexes. pH-2^d-1 and pH-2^d-3 were digested to completion with two sets of restriction enzymes inactivating the tetracycline resistance (*Tc*^{R}) gene. Neither molecule alone could, in principle, confer Tc^R upon transformation, but heteroduplexes could, provided that the two single strand gaps are repaired (Fig. 1).

In control experiments, pBR322 was cut by one or the other pair of enzymes (*Eco*RI/*Hin*dIII or *Bam*HI/*Sph* I). When digested molecules of only one type were denatured and reannealed, none or few Tc^R transformants were obtained. When heteroduplexes of pBR322 digested by one and the other set of enzymes were made, about 1–2 × 10^5 Tc^R transformants per μg of DNA were obtained—i.e., 1/10th to 1/20th the number obtained with undigested pBR322 in a *recA*$^-$ or the isogenic *recA*$^+$ host (Table 1).

Heteroduplexes of pH-2^d-1 and pH-2^d-3 were prepared. The transformation efficiency was further decreased (Table 1). However, the number of Tc^R transformants was much higher than (10-fold or more) the backgrounds obtained with self-annealed pH-2^d-1 or pH-2^d-3, or a mixture of nondenatured, or separately self-annealed plasmids (see legend to Table 1).

Restriction Analysis of Tc^R Transformants. Sixty *recA*$^+$ and 48 *recA*$^-$ *Tc*^{R} transformants were reisolated and plasmid DNA was characterized by restriction mapping using *Bgl* II, *Hin*fI, *Hpa* II, and *Rsa* I, which readily discriminate between the parental molecules (Fig. 2). By this test, the 60 *Tc*^{R} *recA*$^+$ transformants distributed about equally between the two parental types, and no other kind of molecule was found. In contrast, 5 of the 48 plasmids isolated in the *recA*$^-$ host displayed a novel combination of restriction sites; the remaining 43 clones were of the two parental types (Table 1).

To examine the possible involvement of *dam* methylation (2, 3, 7, 8) we introduced pH-2^d-1 and pH-2^d-3 into a *dam*$^-$ host and prepared DNA. The extent of methylation of the G-A-T-C sites was monitored with *Mbo* I, *Sau*3A, and *Dpn* I, which recognize the G-A-T-C sequence in different methylation contexts (26, 27). These controls (not shown) indicated that the sites were essentially all methylated or unmethy-

Table 1. Analysis of Tc^R clones obtained on transformation by heteroduplex DNA

Parental molecules	*dam* methylation	*E. coli* host	No. of Tc^R transformants*	No. of clones analyzed	Clone type†			Clones studied
					pH-2^d-1	pH-2^d-3	Rearranged	
pBR322/pBR322	+/+	*recA*$^+$	1.7 × 10^5					
	+/+	*recA*$^-$	1 × 10^5					
pH-2^d-1/pH-2^d-3	+/+	*recA*$^+$	7.4 × 10^3	60	30	29	0	
	+/+	*recA*$^-$	3.8 × 10^3	48	19	24	5	p132, p133 p137, p138
	−/+	*recA*$^-$	5 × 10^3	24	4	18	2	p134
	+/−	*recA*$^-$	8 × 10^3	24	16	7	1	
	−/−	*recA*$^-$	1 × 10^3	20	7	12	1	p136
pH-2^d-3/p133	+/+	*recA*$^-$	8 × 10^3					

*Per microgram of DNA; not normalized with respect to the transformation efficiencies measured with intact pBR322. The latter were about 4 × 10^6 transformants per μg of DNA in the *recA*$^+$ host, and 1.5 × 10^6 in *recA*$^-$, but varied 2- to 3-fold in different experiments. Backgrounds (often 0 and always less than 10^3) have not been subtracted.

†Determined by restriction mapping only.

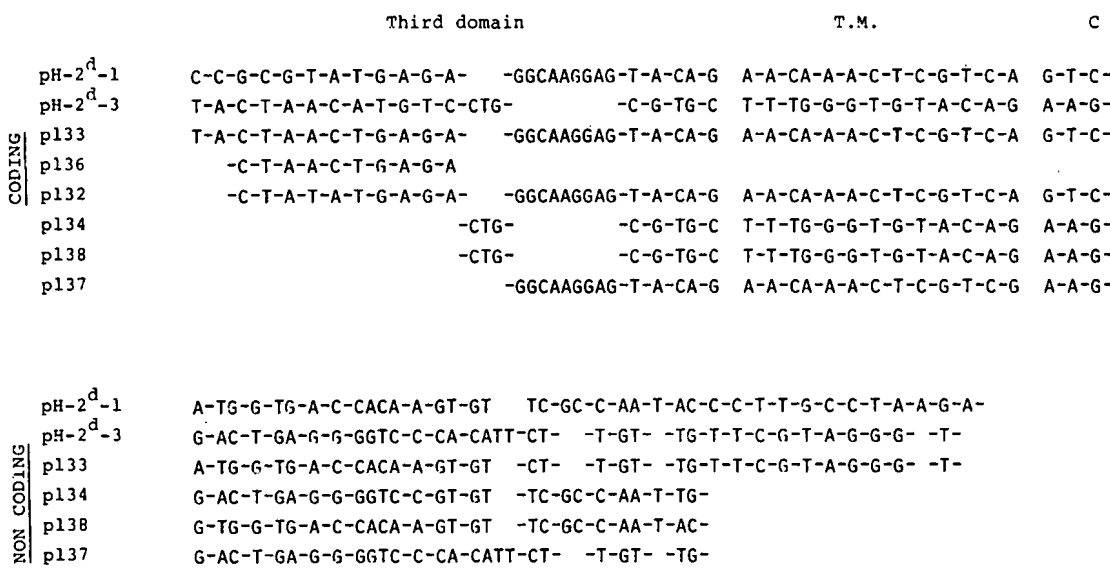


FIG. 3. Nucleotide sequences of rearranged regions. The sequences are arranged with regard to the published sequences of pH-2^d-1 and pH-2^d-3 (20, 21), the latter being corrected for three printing mistakes. For simplicity, the only nucleotides shown are those different in the two plasmids. They are separated by bars, indicating one or several identical nucleotides, or a blank, indicating a deletion with regard to the other aligned sequence. Regions coding for the third extracellular domain, transmembrane (TM), and cytoplasmic (C) parts of the molecule are on top, and the noncoding ones on the bottom. Sequence strategies are as described in Fig. 2.

lated on both strands in plasmids grown in the *dam*$^+$ or *dam*$^-$ host, respectively. Heteroduplexes were then prepared with one methylated and one unmethylated parent, and transformed into *E. coli recA*$^-$. Clones were analyzed as described above. Most were of the methylated parental type (Table 1), indicating that *dam* methylation plays a role. When both parents were unmethylated, fewer transformants were obtained. Several rearranged plasmids emerged from these experiments, two of which were further analyzed (Table 1).

The restriction maps of six plasmids with a rearranged *H-2* sequence are shown in Fig. 2.

Partial Nucleotide Sequence of Rearranged Clones. For these six clones, nucleotide sequences were determined in the region of heterologous restriction sites. In a search for variations undetected by restriction mapping, the entire sequence of one clone (p133) was determined. Data are shown in Fig. 3. Sequences of rearranged clones match that of the two parents without involving any new nucleotide.

DISCUSSION

We have prepared *in vitro* complex heteroduplexes from two sequences differing by more than 8% of their nucleotides and shown that, on transformation and cloning in *E. coli*, rearranged sequences can be obtained. This observation can be accounted for either by two recombination events (between truncated plasmid molecules present in the transformation mixture, or between plasmids generated by replicational segregation of the heteroduplexes) or by heteroduplex repair. In other systems so far studied, the efficiency of repair prevailed over recombination (5, 9). Furthermore, the rearrangements observed here occur apparently at random in a *bona fide recA*$^-$ host. We, therefore, favor heteroduplex repair as the most simple and likely explanation.

Results in Table 1 indicate a bias in favor of the methylated parental sequence when the other one is undermethylated. This could mean that segregants of the parental types are generated through *dam*-directed repair, but it might also reflect increased sensitivity of unmethylated strand in the heteroduplex to nucleolytic action. The latter hypothesis may be more likely, because Pukkila *et al.* (8) have shown that fully methylated heteroduplexes of λ DNA are poorly repaired. The rescue of complex heteroduplexes as parental or rearranged sequences may or may not use presently known repair mechanisms. A variety of *E. coli* mutants deficient in replication, recombination, and repair activities will have to be studied to clarify this question.

In 60 clones isolated on transformation of *recA*$^+$ by heteroduplex DNA, no plasmid displayed a rearranged array of restriction sites. However, further experiments (unpublished observations) show that rearranged clones can be found, but they are 2 to 4 times rarer than in *recA*$^-$. Tc^R transformants are 2 to 3 times more abundant, however. Conceivably, part of the Tc^R transformants obtained in *recA*$^+$ arise by recombination between overlapping truncated plasmids, increasing the background of nonrearranged clones. The figure of ≈10% rearranged clones isolated in *recA*$^-$ (5 out of 48) may be an underestimate, because they were identified by restriction mapping, which may leave alterations undetected.

The summary of our present analysis of six rearranged clones, together with the presumed structure of the initial heteroduplex, is shown in Fig. 4. The structure of the six clones can be interpreted as resulting from a single correction (repair) event in a region either internal to the H-2 cDNA sequences (p133, p134) or overlapping an unknown length of adjacent pBR322 sequence (p132, p136, p137, p138).

The 142 bp present in pH-2^d-1 and absent in pH-2^d-3 must create a large loop in the heteroduplexes. Its correction does not appear to be severely limiting in the production of viable transformants because heteroduplexes that lack it (made of pH-2^d-3 and p133) do not yield many more Tc^R transformants (Table 1). The sequence corresponding to the loop is, however, absent in five of the six clones and may, therefore, be preferentially eliminated. This may not hold for smaller loops because the 9- and 3-nucleotide insertions of pH-2^d-1 and pH-2^d-3 are retained in four and two clones, respectively, out of six.

Nucleotide sequences fitting one or the other parent are shown in Fig. 4. The overlaps correspond to sequences identical in pH-2^d-1 and pH-2^d-3. The lengths of the corrected

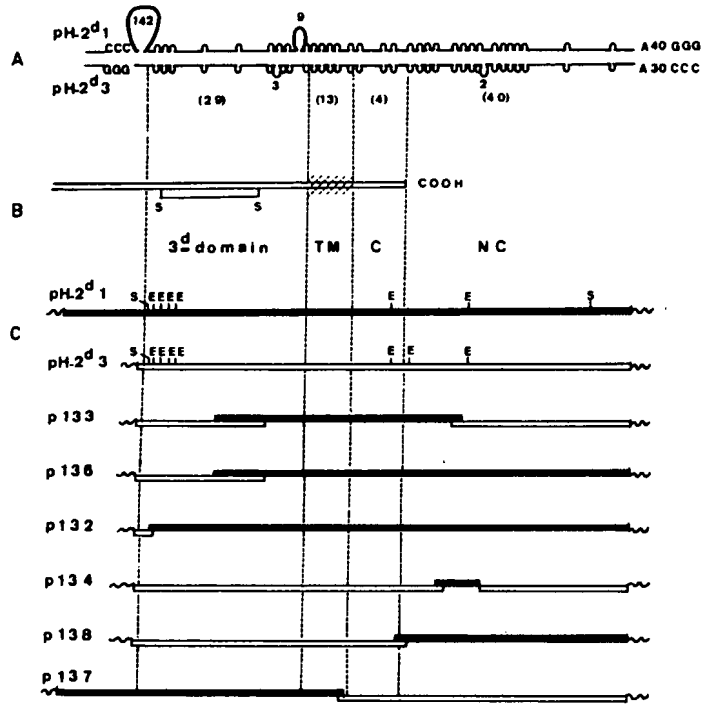


FIG. 4. Structure of rearranged clones. (*A* and *B*) Presumed structure of the starting heteroduplex, showing the 142-nucleotide loop in the 5' region of pH-2^d-1; three smaller loops of 3, 9, and 2 nucleotides; and a number of nucleotide mismatches depicted as small bubbles. Numbers in parentheses indicate the total number of different nucleotides within various regions of the insert illustrated by the COOH-terminal moiety of a H-2 heavy chain (*B*). As in Fig. 3, the third domain, transmembrane (TM), and cytoplasmic (C) coding regions as well as the noncoding region (NC) are indicated. (*C*) The pH-2^d-1 and pH-2^d-3 inserts are shown as filled and open bars, and the *Sau*3A (G-A-T-C) and *Eco*RII (CCA/TGG) sites are indicated as S and E, respectively. The rearranged clones are depicted as filled and open bars as explained in the text.

regions with borders in the overlaps are thus somewhat ambiguous but vary in the approximate range of 75 bp (in p134) to 400 bp (in p133) to 0.5 kb, 0.9 kb, or more in the others. Earlier estimates of the average repair tracts with λ heteroduplexes were in the 2- to 3-kb range (5, 9).

The G-A-T-C and CCA/TGG sequences that undergo methylation in *E. coli* (28) are indicated in Fig. 4. We have found so far no obvious correlation between their location and that of the corrected areas, nor have we identified any evident bias in the choice of substituted bases. Finally, all sequences determined so far (a total of ≈3 kb) fit exactly one or the other parental sequence. In this sense, the correction process, apart from shuffling sequences, does not appear to be grossly mutagenic.

Correction of complex heteroduplexes may be used as a practical means of engineering genetic variants. One of its interesting characteristics is that all features of the primary structure common to both parents are conserved in the variants. Thus, plasmids p132, p133, and p136, which display rearrangements in the coding region, keep the appropriate reading frame and represent mutants of the COOH-terminal half of H-2 molecules. They, of course, retain all usual traits of heavy chains (17) (glycosylation and phosphorylation sites, cysteins in the appropriate position to make a disulfide bridge, etc.).

Many sequences, particularly in the higher eukaryotes, are only partially homologous and differ in many nucleotides. In spite of this, it has often been postulated (3, 16, 29–33) that they can undergo crossing-over and gene conversion on the basis of their (partial) homology. As was emphasized earlier (14), if hybrid DNA is involved in any of these genetic exchanges, it must be in the form of complex heteroduplexes between the partially homologous sequences. Beyond a possible important evolutionary significance (14) the resolution of these complex heteroduplexes into homoduplexes has at least two interesting implications. (*i*) It offers a mutational mechanism capable of altering several nucleotides (amino acids) in a single step, a process that has often been postulated on the basis of amino acid sequence comparisons (34–36). (*ii*) It may be the source of considerable genetic diversity (3, 13, 14) especially if independent correction events generate patchworks (see ref. 14 for elementary calculations on the number of variants generated). In this regard, we proposed in a variant of the mosaic gene model (16) that it might account for at least some of the variations currently attributed to gene conversion (16, 29, 37, 38), which underlie the polymorphism of class I histocompatibility antigens. Indeed, our observation that an H-2 cDNA sequence could be interpreted as a patchwork of two others (29) initially called our attention to heteroduplex correction.

Our results are only valid for *E. coli*, where heteroduplex correction may have played a role—for instance, in the evolution of temperate phage genomes. Whether, as we predict, an extrapolation will hold for genes of the higher eukaryotes requires experimentation in animal cells.

We are grateful to Dr. M. Radman for helpful discussions and to Drs. M. Cochet and J. P. Abastado for advice in Maxam–Gilbert sequence analysis. This work was supported by Institut National de la Santé et de la Recherche Médicale (S.C. 20 and P.R.C. 124031), Centre National de la Recherche Scientifique (E.R. 201 and A.T.P. 90.5039), and the Fondation pour la Recherche Médicale Française.

1. Radding, C. M. (1978) *Annu. Rev. Biochem.* **47**, 847–880.
2. Radman, M., Wagner, R. E., Jr., Glickman, B. W. & Meselson, M. (1980) in *Progress in Environmental Mutagenesis*, ed. Alacevic, M. (Elsevier/North-Holland, New York), pp. 121–130.
3. Bourguigon Van Horen, F., Brotcorn, A., Caillet-Fauquet, P., Diver, W. P., Dohet, C., Doubleday, O. P., Lecomte, P., Maenhaut-Michel, G. & Radman, M. (1982) *Biochimie* **64**, 559–564.
4. Bale, A., d'Alarcao, M. & Marinus, M. G. (1979) *Mutat. Res.* **59**, 157–165.
5. Wildenberg, J. & Meselson, M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2202–2206.
6. White, R. L. & Fox, M. S. (1975) *Mol. Gen. Genet.* **141**, 163–171.
7. Radman, M., Dohet, C., Bourguignon, M. F., Doubleday, O. P. & Lecomte, P. (1981) in *Chromosome Damage and Repair*, eds. Seeberg, E. & Kleppe, K. (Plenum, New York), pp. 431–445.

8. Pukkila, P. J., Peterson, J., Herman, G., Modrich, P. & Meselson, M. (1983) *Genetics* **104,** 571–582.
9. Wagner, R., Jr., & Meselson, M. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 4135–4139.
10. Lai, C. J. & Nathans, D. (1975) *Virology* **66,** 70–81.
11. Miller, L. K., Cooke, B. A. & Fried, M. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3073–3077.
12. Wilson, J. H. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3503–3507.
13. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. (1978) *Science* **202,** 11–17.
14. Kourilsky, P. (1983) *Biochimie* **65,** 85–93.
15. Klein, J. (1978) *Adv. Immunol.* **26,** 55–146.
16. Gachelin, G., Dumas, B., Abastado, J. P., Cami, B., Papamatheakis, J. & Kourilsky, P. (1982) *Ann. Immunol. (Inst. Pasteur)* **133,** 3–20.
17. Kimball, E. S. & Coligan, J. E. (1983) *Contemp. Top. Immunol.* **9,** 1–58.
18. Hood, L., Steinmetz, M. & Malissen, B. (1983) *Annu. Rev. Immunol.* **1,** 529–568.
19. Murray, N. Z., Brammar, W. J. & Murray, K. (1977) *Mol. Gen. Genet.* **150,** 53–61.
20. Kvist, S., Brégégère, F., Rask, L., Cami, B., Garoff, H., Daniel, F., Wiman, K., Larhammar, D., Abastado, J. P., Gachelin, G., Peterson, P. A., Dobberstein, B. & Kourilsky, P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2772–2776.
21. Brégégère, F., Abastado, J. P., Kvist, S., Rask, L., Lalanne, J. L., Garoff, H., Cami, B., Wiman, K., Larhammar, D., Peterson, P. A., Gachelin, G., Kourilsky, P. & Dobberstein, B. (1981) *Nature (London)* **292,** 78–81.
22. Kalderon, D., Ooska, B. A., Ely, B. K. & Smith, A. E. (1982) *Nucleic Acids Res.* **10,** 5161–5181.
23. Dagert, M. & Erlich, S. D. (1979) *Gene* **5,** 23–28.
24. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
25. Roychoudhyry, R. & Wu, R. (1980) *Methods Enzymol.* **65,** 43–47.
26. Gelinas, R. E., Myers, P. A. & Roberts, R. J. (1977) *J. Mol. Biol.* **114,** 169–180.
27. Geier, G. & Madrich, P. (1979) *J. Biol. Chem.* **254,** 1408–1413.
28. McClelland, M. (1981) *Nucleic Acids Res.* **9,** 5859–5866.
29. Lalanne, J. L., Brégégère, F., Delarbre, C., Abastado, J. P., Gachelin, G. & Kourilsky, P. (1982) *Nucleic Acids Res.* **10,** 1039–1049.
30. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21,** 627–638.
31. Miyata, T., Yasumaya, T., Yamawaki-Katooka, Y., Obata, M. & Honjo, T. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2143–2147.
32. Baltimore, D. (1981) *Cell* **24,** 592–594.
33. Egel, R. (1981) *Nature (London)* **290,** 191–192.
34. Yeh, W. K. & Ornston, L. N. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5365–5369.
35. Yeh, W. K., Shih, C. & Ornston, L. N. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3794–3797.
36. Lopez de Castro, J. A., Strominger, J. L., Strong, D. M. & Orr, H. T. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3813–3817.
37. Schulze, D. J., Pease, L. R., Geier, S. S., Reyes, A. A., Sarmiento, L. A., Wallace, R. B. & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2008–2011.
38. Weiss, E. H., Mellor, A., Golden, L., Fahener, K., Simpson, E., Hurst, J. & Flavell, R. A. (1983) *Nature (London)* **301,** 671–674.
39. Bastia, D. (1978) *J. Mol. Biol.* **124,** 601–639.
40. Rougeon, F., Kourilsky, P. & Mach, B. (1975) *Nucleic Acids Res.* **2,** 2365–2378.
41. Birnbaum, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.

Gene, 29 (1984) 255–261
Elsevier

GENE 1036

Recombination following transformation of *Escherichia coli* by heteroduplex plasmid DNA molecules

(Recombinant DNA; genetic crosses; in vivo DNA repair)

Shing Chang, Diana Ho, Jane R. McLaughlin and Sheng-Yung Chang

Cetus Corporation, 1400 Fifty-Third Street, Emeryville, CA 94608 (U.S.A.) Tel. (415) 420-3300

(Received February 10th, 1984)
(Revision received and accepted April 16th, 1984)

SUMMARY

Circular heteroduplex DNA molecules introduced into *Escherichia coli*-competent cells are converted to new recombinant plasmids as a result of enzymatic actions in vivo. A pair of plasmids with partial sequence homology were each linearized at a different position with restriction enzymes, and the termini were made flush with the single-strand-specific S1 nuclease. Duplex molecules were then formed by melting and annealing these plasmid DNAs together. In contrast to linear homoduplex molecules, heteroduplexes circularize and therefore transform *E. coli* efficiently. Unique DNA sequences on each of the parental strands in the transforming heteroduplexes can be selectively incorporated or deleted as a result of in vivo enzymatic activities in transformed cells. This method permits the generation of new recombinant sequences in vivo without relying solely on the presence of convenient restriction sites for manipulation of DNA fragments in vitro.

INTRODUCTION

Since the advent of recombinant DNA technology, it has been possible to create recombinant plasmids by inserting specific restriction fragments into vector molecules (for a recent review, see Morrow, 1979). The cloned DNA sequence can be further modified enzymatically, or by a number of recently developed methods such as primer-directed mutagenesis (Smith and Gillam, 1981) to produce specific sequence alterations. Therefore, a large number of mutational alleles of any cloned gene can be obtained by using these methods. For further genetic analyses, it is frequently desirable to perform genetic crosses between different alleles. The conventional recombinant DNA methods can be used to replace a segment of cloned DNA containing the sequence of interest with the comparable region containing a different allele. However, this approach requires the presence of convenient restriction endonuclease recognition sites flanking the sequence of interest. We describe here a method that allows efficient genetic recombination between alleles on different plasmids with much less dependence upon specific restriction site placement. This method involves the use of a pair of parental plasmids that share partial sequence homology, one or both of which carries a desirable sequence or mutational allele to be included in a new recombinant progeny plasmid. First, the two parental plasmid molecules are each linearized at a unique but different restriction site. The termini of the mole-

Abbreviations: Ap^R, ampicillin resistant; bp, base pair(s).

cules are then made blunt by S1 endonuclease (Ando, 1966) that has specificity for single-stranded DNA. These two linearized parental plasmids are melted and annealed together to generate both linear homoduplex and circularized heteroduplex molecules. The former cannot transform *E. coli* efficiently, but the latter can because of their circular conformation. The hanging-tails and single-stranded gaps in the heteroduplex molecules can be respectively removed and repaired in transformed cells in vivo, resulting in the formation of new recombinant plasmids. Mispaired sequences due to allelic differences between the parental plasmids can also be incorporated with high frequency into the progeny plasmids via this 'heteroduplex-transformation' procedure. Similar heteroduplex intermediates have been utilized for deletion loop mutagenesis (Kalderon et al., 1982; Peden and Nathans, 1982). The mechanisms involved in the in vivo maturation of heteroduplex plasmid molecules have also been studied (West et al., 1983).

MATERIALS AND METHODS

(a) Bacteria and plasmids

E. coli K-12 strain CS412 (McLaughlin et a 1982) was used as host. Plasmid transformation wa carried out essentially as described by Cohen et a (1972). Ap^R transformants were selected on plate containing 50 μg/ml of Ap. Plasmid DNA was pr pared either from CsCl gradients (Kupersztoch an Helinski, 1973) or by the small-scale method of Isl Horowicz and Burke (1981).

Two plasmids were employed in this study pDH5501 and pSYC716. Plasmid pDH5501 is de rived from pBR322 (Bolivar et al., 1977; Sutcliff 1978) with a 338-bp insert originating from the *Baci lus licheniformis* penicillinase (*penP*) gene (Kroye and Chang, 1981; Neugebauer et al., 1981). Thi insert is located between the *Cla*I and the *Hin*dII sites of pBR322 (Fig. 1). It contains the *penP* se

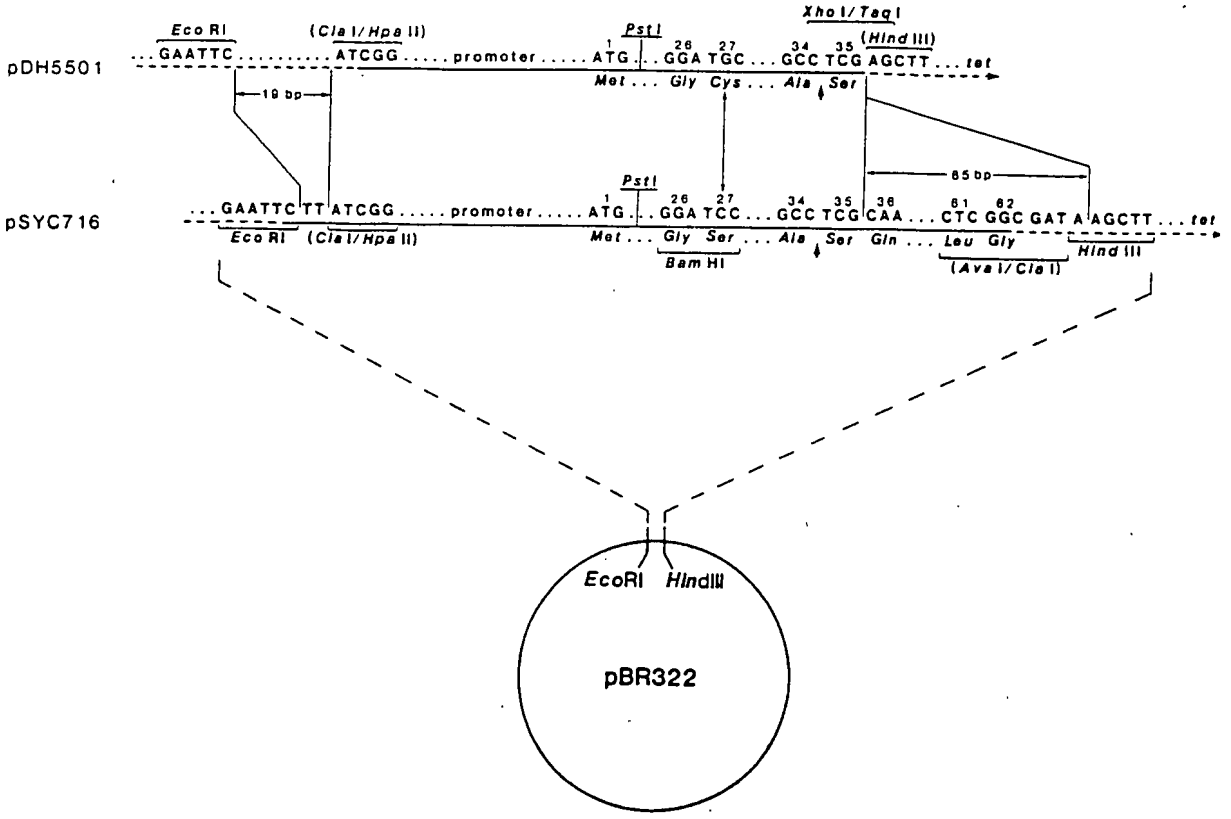


Fig. 1. Sequence differences between plasmids pDH5501 and pSYC716. Both plasmids contain sequences derived from the *penP* gen (solid lines) and from the plasmid pBR322 (dashed lines). The regions containing sequences unique to each plasmid and the S-27 mutatio site (S-27 mutation creates a new *Bam*HI site) are indicated by vertical lines. The locations for certain restriction sites are indicated those shown in parentheses are sites involved in the construction of these plasmids, but they are no longer recognized by these restrictio enzymes due to ligation to heterologous sequences.

quence between the 5′ *Hpa*II site at position 32 according to Neugebauer et al. (1981) and the DNA sequence encoding the 35th amino acid of prepenicillinase, including the *penP* promoter (McLaughlin et al., 1982; Kroyer and Chang, 1981). This amino terminal portion of the penicillinase gene was obtained by controlled removal of nucleotides using BAL31 exonuclease from the *Ava*I site (CTCGGG) corresponding to the coding sequence for the 61st and 62nd amino acid residues (Kroyer and Chang, 1981; McLaughlin et al., 1982; Neugebauer et al., 1981). After a series of manipulations including cleavage by *Eco*RI enzyme, the population of fragments containing various portions of amino terminal coding sequence was cloned into pBR322 between the *Eco*RI and the repaired (described below) *Hin*dIII sites. Plasmids were isolated from transformants and analyzed. One of the plasmids employed for the present study, pDH5501, contains a unique *Xho*I site at the junction (see Fig. 1). This allows convenient manipulation of the promoter-signal peptide sequence of *penP* for expression of heterologous genes and for directed secretion of gene products.

We recently cloned the entire *penP* gene onto the single-stranded M13 phage and carried out primer-directed mutagenesis according to the method of Smith and Gillam (1981). A point mutation (S-27) was introduced into the *penP* gene that generated a *Bam*HI site and converted the 27th codon from UGC (cysteine) to UCC (serine) (Hayashi et al., 1984). This S-27 mutation in the signal sequence, which does not affect the Ap^R phenotype, prevents the conversion of prepenicillinase to the lipoprotein form of mature penicillinase (Lai et al., 1981; Nielsen et al., 1981); it also increases the fraction of the secreted form of mature penicillinase produced by the bacteria. From the M13-*penP* S-27 RF DNA (Hayashi et al., 1984), a 423-bp fragment containing the promoter-signal sequence was purified. This fragment was obtained by *Alu*I (the AGCT recognition sequence is within the *Hin*dIII site) and *Ava*I digestion, and the *Ava*I terminus was repaired in vitro using *E. coli* DNA polymerase I. This DNA fragment was then ligated to plasmid pBR322 previously digested with *Eco*RI and *Cla*I enzymes and repaired by the same procedure. The resulting plasmid, pSYC716, carries this *penP* S-27 sequence (including the *penP* promoter) in place of the *tet* promoter sequence located between the *Eco*RI and the *Cla*I sites in pBR322. The structures of these plasmids are shown in Fig. 1.

(b) Enzymes

T4 DNA ligase was generously provided by D. Gelfand; other enzymes were purchased from New England Biolabs or Bethesda Research Laboratories and were used according to the supplier's specifications.

(c) Heteroduplex preparation

Linearized plasmid DNA was first treated with S1 nuclease (200 U/ml) in buffer containing 300 mM NaCl, 60 mM $ZnSO_4$, and 50 mM Na · acetate, pH 4.6, at a DNA concentration of 50 μg/ml. After 30 min at 22°C, DNA was extracted with phenol and precipitated with ethanol. The DNA pellet was then resuspended in annealing buffer containing 20 mM Tris, pH 7.5, 100 mM NaCl, and 0.5 mM EDTA at 50 μg/ml of DNA in a 1.5-ml eppendorf tube. The solution containing both parental DNAs was heated for 1 min in a boiling water bath and then transferred to a 90°C bath. After gradual cooling to 30°C (in about 4 h), the annealed DNA was either used immediately or stored at 4°C before being used for transformation experiments.

RESULTS

(a) Plasmid constructions

To generate a new plasmid that contains both the desirable S-27 signal peptide mutation (as in pSYC716) and the convenient *Xho*I site that is located 22 bp downstream (as in pDH5501), we designed a simple method that yields the desirable recombinant plasmid by recombination in vivo via transformation of heteroduplex DNA. Plasmids pSYC716 and pDH5501 were digested with *Hin*dIII and *Eco*RI enzymes, respectively, which linearized each of these plasmids at a unique site. These DNAs were further treated with the single-strand-specific S1 endonuclease to generate blunt-ended termini. These linearized plasmid DNAs were mixed in an equimolar ratio, denatured and annealed to form du-

plexes. Both homoduplexes of the parental types and heteroduplex forms were expected. Since homoduplex forms cannot recircularize, they were not expected to transform *E. coli*-competent cells efficiently. This was determined by measuring the transformation efficiency of linearized and S1 nuclease-treated pSYC716 DNA, which showed a 100-fold reduction as compared to heteroduplex forms (not shown). Heteroduplex molecules can circularize into the form shown in Fig. 2A. The

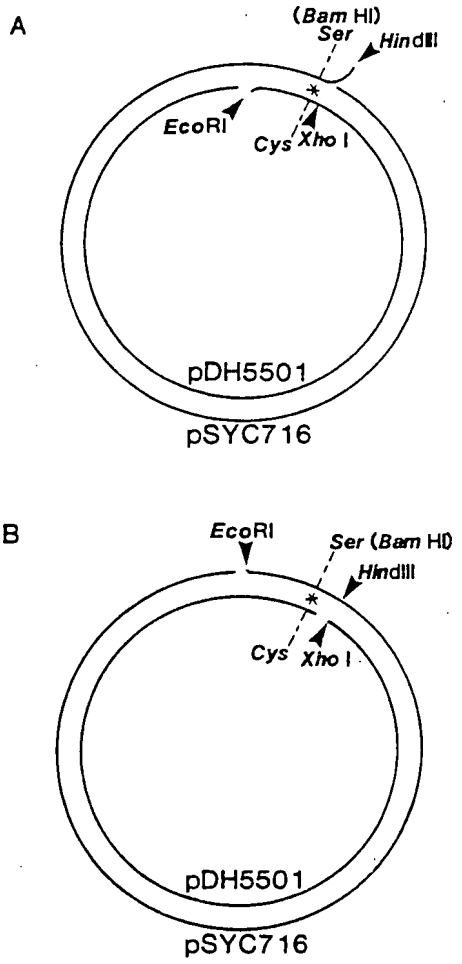


Fig. 2. Expected structures of heteroduplex molecules. (A) Heteroduplex formed between pDH5501 linearized at the *Eco*RI site and pSYC716 linearized at the *Hin*dIII site. Two species of heteroduplex exist due to the two different strands from each homoduplex. In one species the single-stranded tails will be 5′ ends, while in the other species, they will be 3′ ends. (B) Heteroduplex formed between pDH5501 linearized at the *Xho*I site and pSYC716 linearized at the *Eco*RI site. Again, two different species of heteroduplex molecules exist corresponding to those described in (A). See Fig. 1 for detailed sequence comparison between these two plasmids.

single-stranded tails, either as 3′- or 5′-protrudin ends, can be removed by nucleases and the sma single-stranded gaps remaining can be sealed b polymerase and ligase activities in vivo after trans formation. The Ap^R transformants will carry proge ny plasmids of one size class only, which is expecte to be smaller than both parents due to selectiv removal of these sequences.

(b) Transformation by heteroduplexes

Using 0.75 μg each of the two parental plasmid to prepare heteroduplexes for transformation, w obtained 6×10^3 Ap^R transformants. Plasmids fro 16 clones were characterized further; they were a recombinant plasmids of the expected size as revea ed by gel analysis (not shown). Among these 1 clones, one carried a second plasmid indistinguish able from the parental plasmid pSYC716. This prob ably resulted from the cotransformation of pSYC71 that escaped the S1 nuclease inactivation in the het roduplex preparation. *All* of the 16 recombinant pro geny plasmids contained an *Xho*I site inherited fro pDH5501 (not shown). Six of the 16 clones analyze also contained an additional *Bam*HI site, corr sponding to the S-27 mutation in pSYC716. Thi was close to the expected ratio of 8 in 16 assumin each allele has the same probability of being incorpo rated. These data suggested that efficient recomb nation between sequences of pSYC716 an pDH5501 had occurred. One representative clone c each of the two types of progeny plasmids (plasmid pJM86 and pJM106 which carry the S-27 and th wild-type alleles, respectively) was analyzed by re striction digestions, as shown in Fig. 3.

To further investigate the ability of heteroduple DNA to transform and recombine in *E. coli*, we li nearized pSYC716 and pDH5501 at the *Eco*RI an *Xho*I sites, respectively, and converted the termini c the digested DNA to blunt-ends as before. Heterc duplexes were prepared and transformed into *E. col* A total of 1.1×10^4 Ap^R transformants were obtain ed from the heteroduplex molecules prepared fror 2 μg of each of the parental DNA. Plasmids from 2 clones were characterized by restriction analysis (nc shown). They were all the same size and larger tha the parents. This was expected if both of the uniqu regions in the parental plasmids were incorporate into the progeny.

(c) Analysis of heteroduplex transformants

Fig. 2B shows the expected form of the heteroduplex in this experiment. The single-stranded gaps were repaired in vivo which resulted in the incorporation of these sequences into progeny plasmids. Of the 24 clones characterized, only one carried a single *Bam*HI site as in pDH5501: 23 out of 24 carried two *Bam*HI sites, indicating the S-27 allele was preferentially incorporated in this experiment. Because the wild-type allele is located only 22 bases away from one of the two gap regions, exonuclease actions in vivo probably excised this mismatched sequence at the S-27 locus on the duplex DNA before polymerase activity sealed the single-stranded gap region (using the DNA strand containing the S-27 allele as template). In contrast, the S-27 allele is located 324 bases from the other gap and is apparently protected from exonucleolytic removal. Representative progeny plasmids that carry the wild-type (pDH5619) and the S-27 alleles (pDH5618) were analyzed by restriction digestions as shown in Fig. 3. The data confirmed the predicted recombinational events between pSYC716 and pDH5501.

DISCUSSION

A simple method is described that permits efficient and selective transformation of heteroduplex DNA molecules to generate recombinants. Using S1 nuclease to create flush-ended termini on linearized plasmids, it is possible to suppress the background transformation of homoduplex molecules. We also used the repair reaction of DNA polymerase I to generate flush ends and obtained similar results (Chang, S.-Y., unpublished). The data presented here show that single-stranded hanging tails in heteroduplex molecules are selectively and efficiently removed and single-stranded gaps are repaired in vivo in transformed *E. coli* cells. Furthermore, different alleles from the parental plasmids can be introduced into the progeny plasmids with high frequency. This manipulation allows the generation of new recombinant plasmids without total dependence upon the presence of convenient restriction sites flanking the region of interest.

In our experimental design, each pair of linearized parental plasmids can yield two forms of heteroduplex molecules (see legend Fig. 2). In the two cases illustrated in Fig. 2A and 2B, we were unable to

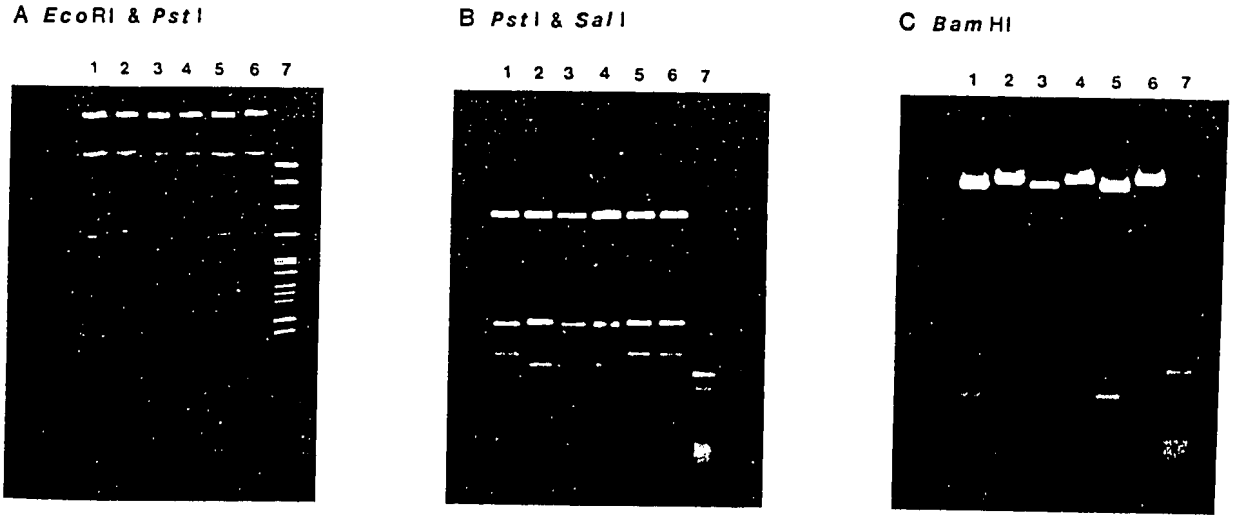


Fig. 3. Gel analyses of restricted fragments from the parental and progeny plasmids. Parental plasmids pSYC716 (1), pDH5501 (2), and the progeny plasmids pJM86 (3) and pJM106 (4) (resulted from heteroduplexes shown in Fig. 2A), and pDH5618 (5) and pDH5619 (6) (resulted from heteroduplexes shown in Fig. 2B), were digested with restriction enzymes as indicated. Plasmids pJM86 and pDH5618 carry the S-27 allele of *penP*, whereas pJM106 and pDH5619 carry the wild-type allele as revealed by the presence or the absence of a second *Bam*HI site in them (panel C). The removal (lanes 3 and 4) or repair (lanes 5 and 6) of single-stranded regions located between the *Eco*RI-*Pst*I and the *Pst*I-*Sal*I sites are shown in panels A and B, respectively. The fragments were resolved on a 4% polyacrylamide gel for (A) and 1.2% agarose gels for (B) and (C). M_r markers in lane (7) were *Hpa*II-digested pBR322 DNA.

distinguish whether only one or both forms were capable of transforming *E. coli* with the same high efficiency. This is because progeny plasmids derived from either one of the two forms of transforming heteroduplex will yield the same expected structure. In both experiments we observed the presence of only one type of progeny plasmid from each characterized transformant with regard to the presence or absence of the S-27 allele; that is, each transformant harbored a plasmid with either the wild-type allele or the S-27 mutant allele on it, but not both types. Since Ap^R transformants were directly picked from the plates, and plasmids were isolated and characterized from these cells without further purification, this suggested that prior to the completion of the first round of DNA replication of the transforming plasmid, the mispaired sequence at the S-27 locus had already been corrected in the transformed cells. In the first experiment (Fig. 2A), each template was used with almost equal frequency for correcting the mispaired sequence; whereas in the second experiment (Fig. 2B), the S-27 mutation was preferentially copied and incorporated. This could be due to the close proximity (22 bases, see Fig. 1) between the wild-type allele and the single-stranded gap shown in Fig. 2B, which made it susceptible to nucleolytic removal prior to polymerase repair in vivo.

We refer to the generation of progeny plasmids in these experiments as recombination due to the appearance of combined traits that were not found together in either of the parents (e.g., the S-27 mutation and the *Xho*I site). This does not imply that the mechanism of this process involves breakage and reunion of double-stranded DNA. The fact that only a single class of recombinant plasmid was observed from each transformant suggests that the progeny recombinant plasmids are generated primarily from efficient repair of mismatched sequences in heteroduplexes, rather than from a crossing-over event between parental plasmid molecules. Whitehouse (1964) was first to propose that heteroduplex structures containing mismatched base pairs could be converted to homoduplex structures by a bacterial error-correcting system(s). More recently, mismatch repair at heteroduplexed regions in plasmid molecules has been reported by West et al. (1983). Thus, by selecting the cleavage sites to generate first linear parental plasmids and then heteroduplexes in vitro, we have demonstrated that the bacterial error-correcting systems can be utilized to maximize the yi of desirable recombinant plasmids in vivo.

In each of the two experiments reported here, linearized the parental plasmids at a site near the e of the mismatched sequences (see Fig. 1), which rected this unique sequence to be located in the he roduplex either as a single-stranded gap or tail. T ing advantage of the abilities of *E. coli* cells to remc hanging tails by nucleases and to repair sing stranded gaps by polymerase activities in vivo, spe fic recombinant plasmids containing either added deleted sequences were obtained from the transf mants. The method described here extends our ab ty to cross DNA sequences or mutations for gener ing new recombinant plasmids without relying up the presence of convenient restriction sites. Sin most cells have enzymatic activities for DNA rep and ligation, this method should work in hosts oth than *E. coli*.

ACKNOWLEDGEMENTS

We are grateful for Dr. David Gelfand's interest this work and for his critical comments on the ma uscript. We also thank Elizabeth Jarvis for editir assistance.

REFERENCES

Ando, T.: A nuclease specific for heat-denatured DNA isolate from a product of *Aspergillus oryzae*. Biochim. Biophys. Act 114 (1966) 158–168.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. Construction and characterization of new cloning vehicle: II. A multipurpose cloning system. Gene 2 (1977) 95–113.

Cohen, S.N., Chang, A.C.Y. and Hsu, L.: Nonchromosomal an tibiotic resistance in bacteria: genetic transformation of *Es cherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. US 69 (1972) 2110–2114.

Hayashi, S., Chang, S.-Y., Chang, S. and Wu, H.C.: Modificatio and processing of *Bacillus licheniformis* prepenicillinase i *Escherichia coli:* fate of mutant penicillinase lacking lipopro tein modification site. J. Biol. Chem. 259 (1984) (in press).

Ish-Horowicz, D. and Burke, J.F.: Rapid and efficient cosmic cloning. Nucl. Acids Res. 9 (1981) 2989–2998.

Kalderon, D., Oostra, B.A., Ely, B.K. and Smith, A.E.: Deletior loop mutagenesis: a novel method for the construction o

point mutations using deletion mutants. Nucl. Acids Res. 10 (1982) 5161–5171.

Kroyer, J. and Chang, S.: The promoter-proximal region of the *Bacillus licheniformis* penicillinase gene: nucleotide sequence and predicted leader peptide sequence. Gene 15 (1981) 343–347.

Kupersztoch, Y.M. and Helinski, D.R.: A catenated DNA molecule as an intermediate in the replication of the resistance transfer factor R6K in *Escherichia coli*. Biochem. Biophys. Res. Commun. 54 (1973) 1451–1459.

Lai, J.-S., Sarvas, M., Brammar, W.J., Neugebauer, K. and Wu, H.C.: *Bacillus licheniformis* penicillinase synthesized in *Escherichia coli* contains covalently linked fatty acid and glyceride. Proc. Natl. Acad. Sci. USA 78 (1981) 3506–3510.

McLaughlin, J.R., Chang, S.-Y. and Chang, S.: Transcriptional analyses of the *Bacillus licheniformis penP* gene. Nucl. Acids Res. 10 (1982) 3905–3919.

Morrow, J.F.: Recombinant DNA techniques, in Wu, R. (Ed.) Methods in Enzymology, Vol. 68, Academic Press, New York, 1979, pp. 3–24.

Neugebauer, K., Sprengel, R. and Schaller, H.: Penicillinase from *Bacillus licheniformis*: nucleotide sequence of the gene and implications for the biosynthesis of a secretory protein in a Gram-positive bacterium. Nucl. Acids Res. 9 (1981) 2577–2588.

Nielsen, J.B.K., Caulfield, M.P. and Lampen, J.O.: Lipoprotein nature of *Bacillus licheniformis* membrane penicillinase. Proc. Natl. Acad. Sci. USA 78 (1981) 3511–3515.

Peden, K.W.C. and Nathans, D.: Local mutagenesis within deletion loops of DNA heteroduplexes. Proc. Natl. Acad. Sci. USA 79 (1982) 7214–7217.

Smith, M. and Gillam, S.: Constructed mutants using synthetic oligodeoxyribonucleotides as site-specific mutagens, in Setlow, J.K. and Hollaender, A. (Eds.) Genetic Engineering, Vol. 3. Plenum, New York, 1981, pp. 1–32.

Sutcliffe, J.G.: pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. Nucl. Acids Res. 5 (1978) 2721–2728.

West, S.C., Countryman, J.K. and Howard-Flanders, P.: Enzymatic formation of biparental figure-eight molecules from plasmid DNA and their resolution in *E. coli*. Cell 32 (1983) 817–829.

Whitehouse, H.L.K.: A theory of crossing-over and gene conversion involving hybrid DNA, in Geerts, S.J. (Ed.) Genetics Today. Proc. 11th Int. Cong., Pergamon, Oxford, 1963, pp. 87–88.

Communicated by J. Hoch.

JOURNAL OF BACTERIOLOGY, June 1989, p. 3046–3052
0021-9193/89/063046-07$02.00/0

Vol. 171, No. 6

Gene Conversion in *Escherichia coli*: the RecF Pathway for Resolution of Heteroduplex DNA

RICHARD FISHEL[1]* AND RICHARD KOLODNER[2]

Laboratory of Chromosome Biology, BRI-Basic Research Program, NCI-Frederick Cancer Research Facility, Frederick, Maryland 21701,[1] and Department of Biological Chemistry, Harvard Medical School, and Laboratory of Molecular Genetics, Dana-Farber Cancer Institute, Boston, Massachusetts 02115[2]

Received 17 October 1988/Accepted 18 March 1989

The independent repair of mismatched nucleotides present in heteroduplex DNA has been used to explain gene conversion and map expansion after general genetic recombination. We have constructed and purified heteroduplex plasmid DNAs that contain heteroallelic 10-base-pair insertion-deletion mismatches. These DNA substrates are similar in structure to the heteroduplex DNA intermediates that have been proposed to be produced during the genetic recombination of plasmids. These DNA substrates were transformed into wild-type and mutant *Escherichia coli* strains, and the fate of the heteroduplex DNA was determined by both restriction mapping and genetic tests. Independent repair events that yielded a wild-type Tet^r gene were observed at a frequency of approximately 1% in both wild-type and *recB recC sbcB* mutant *E. coli* strains. The independent repair of small insertion-deletion-type mismatches separated by 1,243 base pairs was found to be reduced by *recF*, *recJ*, and *ssb* single mutations in an otherwise wild-type genetic background and reduced by *recF*, *recJ*, and *recO* mutations in a *recB recC sbcB* genetic background (the *ssb* mutation was not tested in the latter background). Independent repair of small insertion-deletion-type mismatched nucleotides that were as close as 312 nucleotides apart was observed. There was no apparent bias in favor of the insertion or deletion of mutant sequences.

The repair of, or failure to repair, heteroduplex DNA produced during genetic recombination has been proposed to explain gene conversion and map expansion in bacteria and fungi (8, 14, 16, 27, 33, 35, 40, 43). An analysis of the products of plasmid recombination events in *Escherichia coli* has suggested the frequent formation of regions of symmetric heteroduplex DNA, followed by either co-repair of heteroallelic markers contained in the heteroduplex region or DNA replication (6). These results are consistent with the observation that most mismatch repair events in *E. coli* are characterized by excision-resynthesis tracts that generally cover several thousand nucleotides and would, therefore, lead to co-repair of closely linked mismatched sites (10, 13). However, a proportion of the products observed after genetic recombination could most easily be explained by the formation of symmetric regions of heteroduplex DNA, followed by independent repair of mismatch nucleotides (6), which suggests the existence of several different modes of heteroduplex DNA recognition and repair after genetic recombination.

Several investigators have demonstrated the involvement of the *mutH*, *mutL*, *mutS*, *uvrD* (*uvrE*, *recL*, and *mutU*), and *dam* (DNA adenine methylase) genes in mismatch correction in vivo and in vitro (5, 28, 36, 39). These genes appear to be part of a repair system that is responsible for counteracting the infidelity of DNA replication and act by catalyzing the repair of mismatched nucleotides that were erroneously inserted during DNA replication (31, 37). Repair correction is accomplished by specific excision-resynthesis of the newly replicated, undermethylated DNA strand (25). This system has been designated the Dam-instructed repair pathway because the observed DNA methylation asymmetries that are responsible for the directionality of the repair process are a result of undermethylation of *dam* methylase recognition sites (15, 20, 34). In addition, the *mutS* and *uvrD* gene products have been shown to act in a *dam* methylation-independent mismatch repair reaction (12, 13, 38, 39). Previous studies have shown that the repair of a 10-base-pair (bp) insertion-deletion mismatch by the Dam-instructed pathway appears identical to repair of a single-base mismatch with regard to both genetic requirements for repair and physical characteristics of the repair reaction (13, 42, 44). These small insertion-deletion mismatches appear to be repaired by a mechanism that is different from that of larger (~100 nucleotides) insertion-deletion mismatches, since in the latter case there is a bias toward cleavage of the single-strand loop as well as alternative genetic requirements (38). The processes of repair by the *dam* methylation-dependent and the major *dam* methylation-independent pathways have been shown to lead primarily to co-repair of heteroallelic 10-bp insertion-deletion-type mismatches that are separated by up to 1,243 bp (13, 42, 44). A co-repair reaction would not contribute to the formation of either wild-type or double-mutant recombinants in crosses between closely linked mutations in which both mutant sites are likely to be included in the same heteroduplex region. Such intermediates would be processed by this mismatch repair system to yield a single-mutant parental configuration. Since crosses between closely linked markers are known to yield wild-type and double-mutant recombinants, it appeared possible that a mismatch repair system that could repair closely linked mismatched sites independently of each other might exist. Studies on the effects of *mutHLS* and *uvrD* mutations on general recombination have shown that they either have no effect on heteroallelic crosses or cause a hype-Rec phenotype (7, 13). These results suggested that another mismatch repair system might function during recombination and that, in some cases, the *mutHLS uvrD*-dependent system might interfere with this repair reaction.

* Corresponding author.

TABLE 1. Strains and plasmids

Strain or plasmid	Genotype	Source
E. coli strains		
AB1157	F⁻ *thr-1 leuB6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1 kdgK51 proA2 his-4 argE3 rpsL31 tsx-33 supE44* λ⁻ λˢ	C. C. Richardson
JC2924	Same as AB1157 but also *recA56*	A. J. Clark
JC5519	Same as AB1157 but also *recB21 recC22*	A. J. Clark
JC9239	Same as AB1157 but also *recF143*	A. J. Clark
JC13031	Same as AB1157 but also *recJ153*	A. J. Clark
RDK1540	Same as AB1157 but also *recN1502*::Tn*5*	This laboratory
RDK1541	Same as AB1157 but also *recO1504*::Tn*5*	This laboratory
RDK1542	Same as AB1157 but also *ruvB9 kdgK*⁺	This laboratory
RDK1309	Same as AB1157 but also *ssb-113*	This laboratory
JC7623	Same as AB1157 but also *recB21 recC22 sbcB15*	A. J. Clark
JC3881	Same as JC7623 but also *recF143*	A. J. Clark
JC7606	Same as JC7623 but also *recJ153*	A. J. Clark
RDK1561	Same as JC7623 but also *recN1502*::Tn*5*	This laboratory
RDK1563	Same as JC7623 but also *recO1504*::Tn*5*	This laboratory
Plasmids		
pRDK35	*tet-10* Apʳ	This laboratory
pRDK36	*tet-11* Apʳ	This laboratory
pRDK37	*tet-12* Apʳ	This laboratory
pRDK39	*tet-14* Apʳ	This laboratory

In addition, other studies using substrates containing closely linked mismatch sites have shown *dam* methylation, *mutHLS*- and *uvrD*-independent repair of heteroduplex plasmid and M13mp2 DNA (1, 10, 13, 39) and *dam* methylation, and *mutH*- and *uvrD*-independent repair of heteroduplex bacteriophage λ DNA (38). Lieb has described a very-short-patch mismatch repair system; however, it appears to be specific for a unique class of mismatches related to *dcm* methylation (21).

Mutations in the *recA*, *recF*, *recJ*, *recO*, *ssb*, and *topA* genes of *E. coli* have been found to decrease the frequency of plasmid recombination in addition to affecting several other classes of genetic recombination (2-4, 9, 11, 19, 22-24). The *recA*, *recF*, *recJ*, and *recO* genes, in addition to *recN*, *recO*, and *ruv*, which are not required for plasmid recombination in wild-type *E. coli*, have been collectively designated as belonging to the RecF pathway of genetic recombination in *E. coli* (3, 4, 19, 22-24, 30; C. Luisi-DeLuca, S. T. Lovette, and R. D. Kolodner, Genetics, in press). Aside from *recA* and *ssb*, no definitive function has been found for any of these genes during genetic recombination. One possibility is that one or several genes might affect the repair of mismatched nucleotides produced during genetic recombination, resulting in defective recombination processes.

In this study, we investigated heteroduplex DNA repair in otherwise *rec*⁺ *E. coli* and in an *E. coli* strain containing *recB recC sbcB* mutations in which the RecF pathway genes are activated (3, 4). The study of heteroallelic mismatch repair in these genetic backgrounds was undertaken to evaluate the processing of heteroduplex DNA substrates that mimic putative intermediates in the recombination of plasmids. We previously showed that the independent repair of symmetrically *dam*-methylated heteroduplex plasmid DNA did not require *mutHLS* and *uvrD* gene functions (1, 10, 13). Here we report evidence that the *recF*, *recJ*, *recO*, and *ssb* gene products are involved in the independent repair of small insertion-deletion-type mismatches present in these symmetrically *dam*-methylated heteroduplex plasmid DNAs (13).

MATERIALS AND METHODS

Chemicals and enzymes. Ultrapure Tris was obtained from Schwarz/Mann (Cleveland, Ohio). Restriction endonucleases were obtained from New England BioLabs, Inc. (Beverly, Mass.) and used according to the specifications of the manufacturer. T4 DNA ligase was purified by an unpublished procedure similar to that of Tait et al. (41).

Strains and plasmids. All strains used in this study are isogeneic derivatives of *E. coli* AB1157 (Table 1). All plasmids are derivatives of pBR322 that contain *Xho*I linker insertion mutations in the tetracycline resistance gene and have been described previously (6). pRDK35, -36, -37, and -39 each contain 22 *dam* sites as well as a single 10-bp insertion mutation encoding an *Xho*I cleavage site located at nucleotide positions 24, 339, 652, and 1268, respectively (6). Plasmid DNAs were maintained in and isolated from *E. coli* JC9604 *recA56* (18). These plasmid DNA preparations have been shown to contain less than 1 unmethylated *dam* site per 10 DNA molecules (13).

Construction of heteroduplex plasmid DNA. Heteroduplex plasmid DNA was constructed by using pairs of the pRDK35, pRDK36, pRDK37, and pRDK39 plasmid DNAs originally described by Doherty et al. (6). The construction procedure relies on the observation that restriction endonucleases cleave only double-stranded homoduplex DNA and do not recognize sites that contain mismatched nucleotides (10, 25). Plasmid DNA was linearized by digesting it to completion with *Pst*I. Then it was purified by extraction with phenol, precipitated in ethanol, and suspended in 2 mM EDTA (pH 8.0) at a final DNA concentration of 0.5 to 1 mg/ml. Equal amounts of two plasmid parents were mixed, 2 M NaOH was added to a final concentration of 0.2 M, and the DNA was incubated on ice for 5 min. The solution was then neutralized by addition of an equimolar quantity of 1 M Tris hydrochloride, and the DNA (200 to 500 μg/ml) was reannealed by incubation at 55°C for 15 min. The reannealed DNA was diluted to a final concentration of 1 to 2 μg/ml in 40 mM Tris buffer (pH 7.8)–8 mM $MgCl_2$–5 mM β-mercap-

toethanol–0.067 mM ATP–0.7 U of T4 DNA ligase per ml and incubated at 12.5°C for 6 h. After the ligation step, 0.5 M EDTA (pH 8.0) was added to a final concentration of 10 mM. The DNA was then concentrated to 0.5 ml by ultrafiltration with a YM10 series membrane (Amicon Corp., Lexington, Mass.), extracted with phenol, precipitated with ethanol, and suspended in 2 mM EDTA (pH 8.0) at a final concentration of 200 to 500 μg/ml. The cyclized DNA was then digested to completion with *Xho*I and centrifuged in a solution containing cesium chloride (ρ = 1.55 g/ml), 10 mM EDTA (pH 8.0), and 200 μg of ethidium bromide per ml for 36 h at 40,000 rpm in a 70.1 Ti rotor (Beckman Instruments, Inc., Fullerton, Calif.) at 15°C. The covalently closed circular DNA-containing band was isolated from the equilibrium gradient, dialyzed for 4 h in the dark at 4°C against 1 liter of 10 mM Tris (pH 8.0)–1 mM EDTA, extracted with phenol two times to remove the ethidium bromide, precipitated with ethanol, and resuspended at a final concentration of 100 to 200 μg/ml in 2 mM EDTA (pH 8.0). A typical preparation started with 200 μg of plasmid DNA and yielded approximately 50 μg of homogeneous heteroduplex DNA.

Assay for heteroduplex DNA repair. Transformation of *E. coli* strains with heteroduplex plasmid DNAs was performed by a modification of the method described by Morrison (29). After the calcium and heat shock procedure, the transformation mixture was aerated at 37°C in L broth for 1 h to allow expression of drug resistance. Transformants were plated onto L-broth plates containing ampicillin (50 μg/ml) or ampicillin plus tetracycline (15 μg/ml) to determine the proportion of transformants containing a plasmid that had undergone an independent repair event to produce a wild-type tetracycline resistance gene before replication. The frequencies of transformation of the strains listed in Table 1 by pBR322 derivatives were essentially identical (approximately 2×10^6 to 5×10^6 transformants per μg of DNA) with the exception of *E. coli* JC5519 *recB21 recC22*, which yielded approximately 2×10^5 transformants per μg of DNA. Structural analysis of plasmid DNA isolated from individual transformants was accomplished by rapid plasmid DNA isolation (17) and restriction mapping analysis as described previously (13).

RESULTS

Description of the experimental system. Figure 1 illustrates the steps involved in the preparation of heteroduplex plasmid DNA substrates. We did not purify the individual single strands of pBR322; therefore, the reannealing produced two different configurations of heteroduplex plasmid DNA. It should be noted that the 10-bp *Xho*I linker insertions at the different sites are always located on opposite DNA strands with respect to each other within any single molecule; however, they can occur on either complementary DNA strand with equal probability. We have previously shown that plasmid DNA grown and amplified in *dam*$^+$ *E. coli* contains less than 1 unmethylated *dam* site per 10 molecules and that heteroduplex DNA constructed by using this DNA is symmetrically methylated (13). No parental fragments were observed after double digestion with *Pst*I and *Xho*I (Fig. 1; compare lanes G and K), which indicated that the heteroduplex plasmid DNA preparation had less than 5 ng of homoduplex DNA in a 0.2-μg sample.

There are three possible fates for heteroduplex plasmid DNA introduced into *E. coli* by transformation: (i) DNA replication before repair, (ii) co-repair, and (iii) independent repair (Fig. 2). There are nine product combinations that can

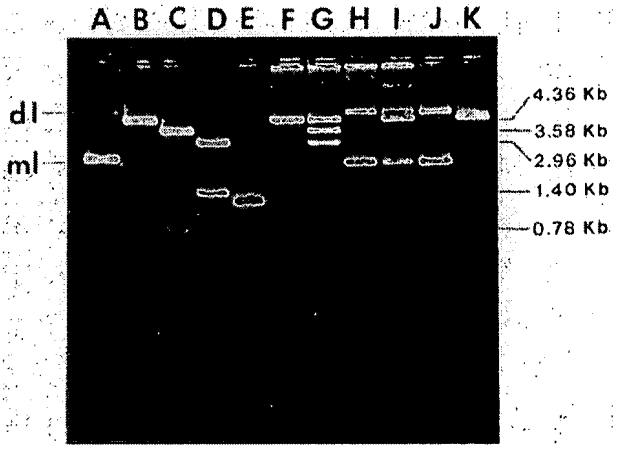


FIG. 1. Analysis of the steps involved in preparation of heteroduplex DNA. Analysis of DNA samples was carried out by electrophoresis through 1% agarose slab gels run in 40 mM Tris (pH 7.9)–5 mM sodium acetate–1 mM EDTA–0.5 μg of ethidium bromide per ml. Lanes: A, undigested mixture of pRDK35 and pRDK37; B, same as lane A except digested with *Pst*I; C, *Pst*I-plus-*Xho*I digest of pRDK35; D, *Pst*I-plus-*Xho*I digest of pRDK37; E, same DNA as in lane B except denatured with 0.2 M NaOH and neutralized with 0.2 M Tris hydrochloride; F, same DNA as analyzed in lane E except after renaturation by incubation at 55°C for 15 min; G, *Xho*I digest of the DNA in lane F (*Xho*I-resistant 4.36-kbp linear heteroduplex DNA); H, DNA analyzed in lane F except cyclized with T4 DNA ligase; I, *Xho*I digest of the DNA analyzed in lane H; J, cesium chloride-ethidium bromide-purified covalently closed circular DNA obtained from the DNA analyzed in lane I; K, *Pst*I-plus-*Xho*I digest of the DNA analyzed in lane J. mI, Covalently closed circular monomer; dI, covalently closed circular dimer.

be distinguished in these experiments. DNA replication will produce two genetically distinct tetracycline-sensitive (Tc^s) plasmid molecules within the same cell, and tetracycline-resistant (Tc^r) plasmid molecules can subsequently be produced by genetic recombination at a rate of 10^{-6} per cell division during growth of the transformant (Luisi-DeLuca et al., in press). Co-repair is the most likely result of an excision-resynthesis repair tract that covers both of the heteroduplex sites (13) and produces a single Tc^s parental plasmid. Single independent repair events followed by replication or two separate independent events, both on a single molecule, account for the remaining products. Tc^r molecules in the transformation mix can be produced only by independent repair or by reversion of a marker site, which has not been observed with these mutations (6, 9). The co-repair of heteroduplex markers separated by 1,243 bp is the most frequently observed repair product (13). Independent repair is distinguished by the formation of a recombinant plasmid that contains sequences derived from both parental single strands that were originally contained in the heteroduplex substrate. The most convenient measure of independent repair for the heteroduplex plasmid substrates described is the detection of Tc^r cells formed immediately after transformation, which is accomplished by plating the transformation mix onto appropriate indicator plates (see Materials and Methods). In addition, independent repair can be detected by analyzing the structure of plasmid DNAs isolated from individual transformants.

***recF*, *recJ*, and *recO* are required for independent repair.** Table 2 shows the frequency of obtaining Tc^r transformants

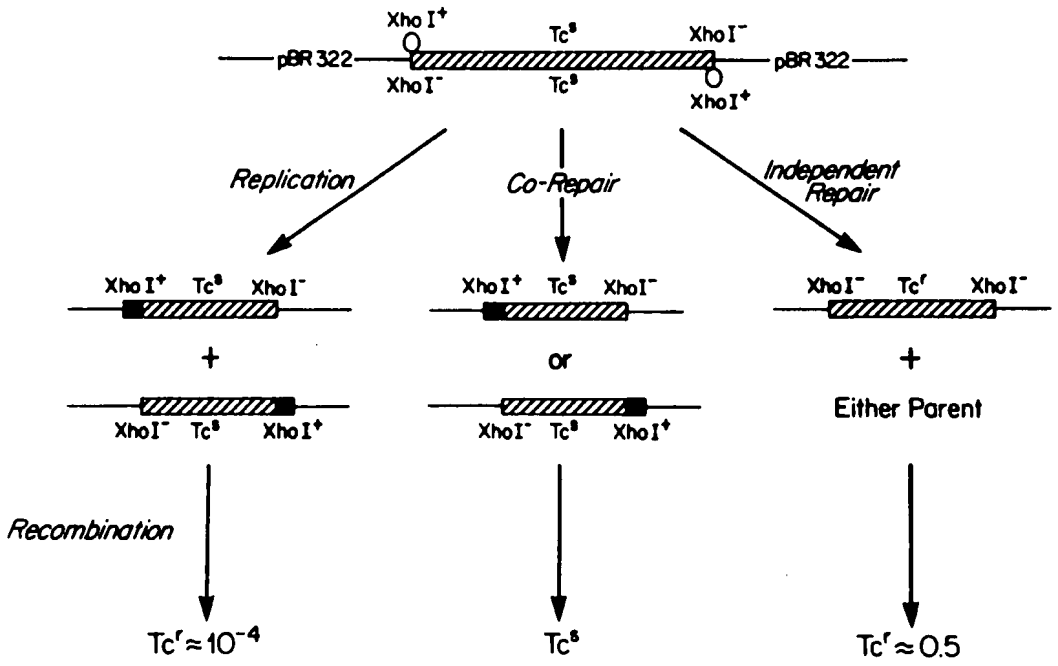


FIG. 2. Possible products formed after transformation of heteroduplex plasmid DNA into *E. coli*. The substrate DNA used contained the illustrated mismatched nucleotides at the sites of two different 10-bp insertion mutations. The insertion mutations were made by inserting *Xho*I linkers at nucleotide positions 24, 339, 652, and 1243 of pBR322 to obtain the *tet-10*, *tet-11*, *tet-12*, and *tet-14* mutations, respectively (6). Replication of this DNA after transformation into an *E. coli* strain will yield the two parental monomeric plasmid DNAs, which will then be present in the transformant colony. Because both parental plasmids are present in such a transformant, Tc^r recombinants will be formed at a frequency of about 10^{-6} per generation during the growth of this class of transformants (Luisi-DeLuca et al., in press). Co-repair, in which a region of DNA covering both mutant sites is excised and resynthesized by using only one strand as the template, will yield only one of the two parental monomeric plasmids, which can then be recovered from the resulting transformant colony. In independent repair, only one of the mismatched sites is repaired, yielding a DNA molecule that still contains one mismatched site. After DNA replication, this molecule will yield two plasmid DNAs; one will be either a wild-type or a double-mutant monomer, and the other will be one of the two parental monomer plasmids. In some cases, two separate, independent repair events can occur to yield a single repair product. Each of the fates described above can then be distinguished by isolating plasmid DNA from the resulting transformants and analyzing its structure by restriction mapping. Simple genetic tests for distinguishing among these possibilities are described in Materials and Methods and reference 13.

after introduction of heteroduplex plasmid DNA into *E. coli* strains containing individual RecF pathway mutations in either a wild-type or a *recB recC sbcB* background. The generation of Tc^r transformants was found to be unaffected by *recA*, *recB recC*, *recN*, and *ruv* mutations regardless of the genetic background. It is unlikely that the formation of Tc^r transformants was due to replication of the substrate followed by recombination (Fig. 2), because plasmid recombination in *E. coli* AB1157 yields Tc^r recombinants at a rate of 10^{-6} per cell division and is *recA* dependent (6, 9, 18, 19; Luisi-DeLuca et al., in press). The yield of Tc^r transformants was reduced 20- to 100-fold when the *recF* or *recJ* mutation was present in an otherwise rec^+ genetic background and in a *recB recC sbcB* genetic background. However, the *recO* gene product was required only for the production of Tc^r transformants in the *recB recC sbcB* genetic background. The *ssb-113* mutation was shown to reduce the frequency of Tc^r transformants eightfold in an otherwise rec^+ genetic background and was not tested in the *recB recC sbcB* background. These results suggest that the events leading to repair of heteroduplex plasmid DNA to yield Tc^r transformants is a complex process requiring different gene products in different genetic backgrounds. The effect of *recF*, *recJ*, *recO*, and *ssb* on heteroduplex plasmid DNA processing is a previously unrecognized phenotype. Mutations in these genes appear to affect repair events that are distinct from the

TABLE 2. Frequency of independent repair of heteroallelic heteroduplex plasmid DNA

Additional host mutation	Frequency of independent repair[a] (%)	
	Wild type[b]	*recB21 recC22 sbcB15*[b]
None	0.88	0.8
recA56	0.57	ND
recB21 recC22	1.6	NA
recF143	0.027	0.028
recJ153	0.013	0.008
recN1502::Tn*5*	0.8	0.7
recO1504::Tn*5*	0.3	0.045
ruvB9	0.62	ND
ssb-113	0.1	ND

[a] Fraction of Tc^r transformants × 100 obtained with pRDK35-pRDK37 heteroduplex DNA. Transformation by individual parental DNAs yielded <0.002% Tc^r transformants. The transformation efficiency for all strains except *recB21 recC22* varied between 2×10^6 and 5×10^6. The average frequency of transformation of *recB21 recC22* was 2×10^5. All frequencies were averaged from at least four independent experiments. *P* values comparing the mutant strains with the parental strains were calculated and considered to be significant if <0.05 (10). All frequencies reported had *P* values of <0.001 except those for *recO*::Tn*5* in an otherwise wild-type background ($P = 0.09$). In these experiments, we considered the results reproducibly significant if frequencies differed by more than eightfold. ND, Not determined; NA, not applicable.

[b] All strains are derivatives of *E. coli* AB1157.

TABLE 3. Effect of distance between heteroduplex DNA sites on independent repair

Relevant host genotype	Frequency of Tc^r transformants obtained (%)[a] with the following no. of base pairs between alleles:			
	313[b]	315[c]	628[d]	1,244[e]
Wild type	0.54	0.51	0.88	0.9
recA56	0.42	0.51	0.57	0.78
recF143	0.02	0.008	0.027	0.036

[a] Determined as described in Table 2, footnote *a*.
[b] Heteroduplex formed between pRDK36 and pRDK37.
[c] Heteroduplex formed between pRDK35 and pRDK36.
[d] Heteroduplex formed between pRDK37 and pRDK39.
[e] Heteroduplex formed between pRDK35 and pRDK39.

repair events catalyzed by the Dam-instructed pathway and from the co-repair reactions observed with methylated substrates (13). In a preliminary study, we have demonstrated that apparent repair of mismatched nucleotides in a plasmid substrate DNA to yield Tc^r molecules can occur regardless of the state of *dam* methylation and is not affected by *mutH*, *mutL*, *mutS*, or *uvrD* mutation (13).

Effect of distance between heteroduplex sites on independent repair. Several reports have demonstrated that mismatch correction in *E. coli* appears to involve excision-resynthesis tracts of greater than 1,000 nucleotides (13, 42). To investigate the effect of the distance between two mismatched sites on the processing of the heteroduplex substrate, we constructed a series of heteroduplex plasmids containing different lengths of DNA between the two insertion-deletion mismatches. Data for the effect of distance between heteroduplex markers on the generation of Tc^r transformants indicated that the distance between the markers had little effect on the frequency production of Tc^r plasmids (Table 3). These results indicated that the class of repair events detected by the heteroduplex DNA processing assay could be explained by repair events involving relatively short (less than 312 nucleotides) excision-resynthesis tracts. The frequency of co-repair (13) was also unaffected by the amount of intervening DNA up to 1,243 nucleotides, which suggested that excision-resynthesis tracts leading to co-repair are greater than this distance (data not shown).

Products of heteroduplex DNA repair. We have shown that it is possible to examine the products of a single heteroduplex repair event by analyzing the structure of the plasmid DNA obtained from a single ampicillin-resistant transformant (13). We carried out such an analysis of the products of repair of a pRKD35-pRDK37 heteroduplex substrate to verify that appropriate DNA products were formed at frequencies consistent with the results obtained with the genetic test for independent repair described above. In addition, this analysis was used to determine whether *recA*, *recF*, and *recJ* mutations have an effect on the frequency of co-repair, since our previously described genetic test for co-repair is not applicable to the analysis of mutations that decrease the frequency of recombination (13). Presented in Fig. 3 is a restriction mapping analysis of plasmid DNA purified from five different transformants that were obtained by transforming *E. coli* AB1157 to Ap^r with heteroduplex plasmid DNA. Results obtained from several experiments in which the structures of the plasmid DNAs obtained from individual transformants were determined (Table 4) indicated that the major product observed was either pRDK35 or pRDK37, which was most likely formed by co-repair as previously discussed (13). The *recA*, *recF*, and *recJ* mutations had no effect on the frequency of co-repair. Ten

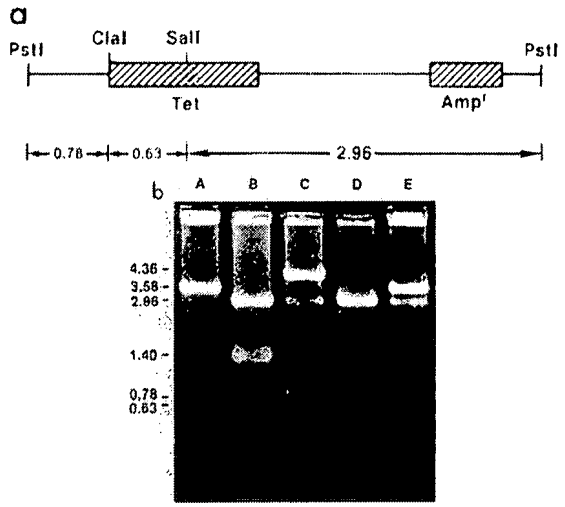


FIG. 3. (a) Structures of heteroduplex DNA repair products as determined by restriction mapping. *E. coli* AB1157 was transformed to Ap^r with pRDK35-pRDK37 heteroduplex plasmid DNA. Structures of the plasmid DNAs present in the transformants were determined essentially as described elsewhere (13) by first verifying that the plasmids were circular monomers and then by digesting the plasmid DNA samples with *Xho*I and *Pst*I to distinguish between the different plasmid DNAs that were present. (b) Identities of plasmid DNAs obtained from five transformants and possible mechanisms by which they were formed. Lanes: A, pRDK35 (*tet-10*) parental plasmid that could have formed by co-repair, yielding 3.58- and 0.78-kbp fragments (10); B, pRDK37 (*tet-12*) parental plasmid that could have been formed by co-repair, yielding 2.96- and 1.40-kbp fragments (10); C, pRDK37 associated with wild-type pBR322 that could have resulted from a single independent repair event followed by DNA replication, yielding 4.63-, 2.96-, and 1.40-kbp fragments (see Fig. 2); D, double-mutant plasmid associated with pRDK35 that could have resulted from a single independent repair event followed by DNA replication, yielding 3.58-, 2.96-, 0.78-, and 0.63-kbp fragments and confirmed in an *Xho*I digest that yielded 0.63-, 3.73-, and 4.36-kbp fragments; E, pRDK35 in association with pRDK37 that could have resulted from replication of the heteroduplex substrate, yielding fragments of 3.58, 2.96, 1.40, and 0.78 kbp.

transformants of the wild-type and *recA* *E. coli* strains contained either wild-type or double-mutant plasmids, which is consistent with the frequency of formation of Tc^r DNA molecules, assuming that they were formed by independent repair (Table 2). In nine of these cases, a single apparently independent repair event appeared to have occurred, since the wild-type or double-mutant plasmid was found in association with a parental plasmid; in one case, two apparently independent repair events may have occurred, since the wild-type plasmid was not found in association with a parental plasmid (Table 4). This latter plasmid configuration could have been generated by a single independent repair event followed by strand loss similar to that observed in λ heteroduplex repair experiments. Although we acknowledge this possibility, it does not appear to be significant for the plasmid repair events described here, since in nearly all of the cases showing independent repair (single-repair events), we can recover all of the DNA strands. Furthermore, as has been shown in other studies on different types of mismatch repair, the repair events observed here require distinct genetic elements known to be required for DNA repair, which provides additional evidence that strand loss does not account for the repair products observed. In all cases

TABLE 4. Restriction mapping analysis of heteroduplex plasmid DNA repair products

Strain tested	Plasmid DNA observed[a]						
	pRDK35	pRDK37	Wild type	Double mutant	pRDK35 + pRDK37	% Co-repair[b]	Plasmid index[c]
Wild type	91	71	3[d]	4[e]	150	51	0.56
recA56	41	37	1[f]	2[g]	39	65	0.53
recF143	11	7	ND	ND	6	75	0.65
recJ153	20	16	ND	ND	24	60	0.56

[a] Plasmid DNA was isolated from individual Ap^r transformants and analyzed by restriction mapping as previously described (10). Numbers shown are actual numbers of transformants present in each category. ND, Not detected, although the sample size was too small to indicate that the product was not formed.
[b] (pRDK35 + pRDK37)/total number of transformants.
[c] pRDK35/(pRDK35 + pRDK37).
[d] Two of the wild-type plasmid products were associated with pRDK37, and the third was not apparently associated with any other plasmid molecule.
[e] Two of the double-mutant plasmid products were associated with pRDK35, and the other two were associated with pRDK37.
[f] Associated with pRDK35.
[g] Both double-mutant plasmids were associated with pRDK35.

examined (four total), the formation of a wild-type Tc^r gene was accompanied by a gain of wild-type restriction sites (in this case, *Cla*I and *Sal*I sites), which eliminates the possibility of some type of aberrant repair event.

DISCUSSION

Previous analysis of mismatch repair in *E. coli* has demonstrated that the major mismatch repair reaction involves excision-resynthesis tracts that are an average of 3 kbp long (5, 13, 25, 28, 37, 42). Consequently, when a heteroduplex DNA molecule contains multiple closely linked mismatched sites, they will most likely be repaired in a concerted event, using the same strand as a template. This will result in a product containing a parental configuration of markers. In other studies, we and others demonstrated that, at a low frequency, multiply marked heteroduplex DNAs appeared to be repaired in a reaction that did not appear to be catalyzed by the major *mutHLS*- and *uvrD*-dependent mismatch repair reaction (1, 10, 13). Furthermore, these events appeared to yield a recombinant configuration of markers (1). In this study, we have analyzed the products formed from heteroduplex plasmid DNAs having two closely linked mismatch sites which contain one mutant site per strand. We found that recombinant configurations were most frequently generated by repair at only one site, followed by DNA replication, which we call independent repair. Such an event generates one parental molecule and one recombinant molecule. The formation of a wild-type Tc^r gene was accompanied by the appearance of the wild-type restriction sites (in this case, *Cla*I and *Sal*I), which reduced the possibility that some aberrant replication error, induced by the palindromic heterologies, was responsible for the Tc^r phenotype (data not shown). A small proportion of the products (<10%) observed appear to be the result of two independent repair events. Independent repair exhibits parity, since wild-type and double-mutant configurations are formed with equal frequency.

Genetic analysis has demonstrated that these apparent repair reactions require *recF*, *recJ*, and *ssb* in an otherwise wild-type genetic background and *recF*, *recJ*, and *recO* in a *recB recC sbcB* genetic background (*ssb* was not tested). Because these studies use 10-bp insertion mismatches, further studies will be required to determine whether single nucleotide substitution mispairs and larger insertion and deletion mispairs can be processed by this reaction. The repair reaction described here appears to involve short excision-resynthesis tracts and to represent a mismatch repair reaction that is distinct from *mutHLS*- and *uvrD*-dependent mismatch repair, from very-short-patch repair (21), and from the recently reported *mutY* system (33) and further supports the idea that *E. coli* contains multiple mechanisms for processing mispaired nucleotides (39).

One of the reasons for analyzing the repair of mismatch-containing plasmid substrates was to gain insight into the role that mismatch repair plays in the recombination of plasmids. Since *E. coli* strains containing *recF*, *recJ*, *recO*, and *ssb* mutations have been shown to be deficient in plasmid recombination, it is tempting to ascribe the single role of these genes in recombination to the repair of mismatched nucleotides after symmetric heteroduplex formation. An analysis of the products of plasmid recombination presented by Doherty et al. (6) suggests that most of the products processed from heteroduplex intermediates are a result of co-repair or segregation after DNA replication and that only a small proportion of the products can have resulted from independent repair. Such an observation is consistent with several studies showing that co-repair, which does not require the *recF* or *recJ* gene product, is more frequent than the independent repair reactions described here (13). An alternative explanation for the role of *recF*, *recJ*, *recO*, and *ssb* in recombination and repair suggests that a RecF pathway reaction introduces short, single-strand gaps or displaced strands into the DNA, possibly at the site of spontaneous DNA damage, and that those structures can subsequently promote genetic recombination. The independent repair of heteroduplex plasmid DNA could result if the same RecF pathway-dependent process could recognize a mismatched site. The concept of heteroduplex DNA repair catalyzed by the RecF pathway was first introduced by Mahajan and Datta (26), who observed multiple independent recombination events catalyzed by the RecF pathway. The introduction of long stretches of single-stranded donor DNA into the recipient chromosome, followed by independent repair of mismatched sites, was proposed as an explanation for these observations. Our results support this idea. It remains unclear whether the *recF*, *recJ*, *recO*, or *ssb* gene product plays a direct or regulatory role in these processes.

ACKNOWLEDGMENTS

We thank A. J. Clark and C. Luisi-DeLuca for many helpful discussions and Janet Andersen for the statistical analysis described in Table 2.

This work was supported by Public Health Service grant GM26017 from the National Institutes of Health and American Cancer Society grant FRA-271 to R.K., by Public Health Service

postdoctoral fellowship GM07693 from the National Institutes of Health to R.F., and in part by the National Cancer Institute, under contract N01-CO-74101 with BRI (R.F.).

LITERATURE CITED

1. **Abastado, J.-P., B. Cami, T. H. Dihn, J. Igolen, and P. Kourilsky.** 1984. Processing of complex heteroduplexes in *Escherichia coli* and Cos-1 monkey cells. Proc. Natl. Acad. Sci. USA **81**:5792–5796.
2. **Clark, A. J.** 1973. Recombination-deficient mutants of *E. coli* and other bacteria. Annu. Rev. Genet. **7**:67–86.
3. **Clark, A. J.** 1980. A view of the recBC and recF pathways of *E. coli* recombination. UCLA/ICN Symp. Mol. Cell. Biol. **18**: 891–899.
4. **Clark, A. J., M. R. Volkert, and L. J. Margossian.** 1979. A role for *recF* in repair of UV damage of DNA. Cold Spring Harbor Symp. Quant. Biol. **43**:887–892.
5. **Claverys, J.-P., and S. A. Lacks.** 1986. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. Microbiol. Rev. **50**:133–165.
6. **Doherty, M. J., P. T. Morrison, and R. Kolodner.** 1983. Genetic recombination of bacterial plasmid DNA: physical and genetic analysis of the products of recombination in *Escherichia coli*. J. Mol. Biol. **167**:539–560.
7. **Feinstein, S. I., and K. B. Low.** 1986. Hyper-recombining recipient strains in bacterial conjugation. Genetics **113**:12–33.
8. **Fincham, J. R. S., and R. Holliday.** 1970. An explanation of fine structure map expansion in terms of excision repair. Mol. Gen. Genet. **109**:309–322.
9. **Fishel, R. A., A. A. James, and R. Kolodner.** 1981. RecA-independent general genetic recombination of plasmids. Nature (London) **294**:184–186.
10. **Fishel, R. A., and R. Kolodner.** 1983. Gene conversion in *Escherichia coli*: the identification of two repair pathways for mismatched nucleotides. UCLA Symp. Mol. Cell. Biol. **11**: 309–326.
11. **Fishel, R. A., and R. Kolodner.** 1984. *Escherichia coli* strains containing mutations in the structural gene for topoisomerase I are recombination deficient. J. Bacteriol. **160**:1168–1170.
12. **Fishel, R. A., and R. Kolodner.** 1984. An *Escherichia coli* cell-free system that catalyzes the repair of symmetrically methylated heteroduplex DNA. Cold Spring Harbor Symp. Quant. Biol. **49**:603–609.
13. **Fishel, R. A., E. Siegel, and R. Kolodner.** 1986. Gene conversion in *Escherichia coli*: resolution of heteroallelic mismatched nucleotides by co-repair. J. Mol. Biol. **188**:147–157.
14. **Fox, M. S.** 1978. Some features of genetic recombination in procaryotes. Annu. Rev. Genet. **12**:47–68.
15. **Herman, G. E., and P. Modrich.** 1982. *Escherichia coli dam*-methylase: physical and catalytic properties of the homogeneous enzyme. J. Biol. Chem. **257**:2605–2612.
16. **Holliday, R.** 1964. A mechanism for gene conversion in fungi. Genet. Res. **5**:282–304.
17. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparations of bacterial plasmids. Anal. Biochem. **141**: 193–196.
18. **James, A. A., P. T. Morrison, and R. Kolodner.** 1982. Genetic recombination of bacterial plasmids: analysis of the effect of recombination-deficient mutations on plasmid recombination. J. Mol. Biol. **160**:411–430.
19. **Kolodner, R., R. A. Fishel, and M. Howard.** 1985. Genetic recombination of bacterial plasmid DNA: effect of RecF pathway mutations on plasmid recombination in *Escherichia coli*. J. Bacteriol. **163**:1060–1066.
20. **Lahue, R. S., S.-S. Su, and P. Modrich.** 1987. Requirement for d(GATC) sequences in *Escherichia coli mutHLS* mismatch correction. Proc. Natl. Acad. Sci. USA **84**:1482–1486.
21. **Lieb, M.** 1983. Specific mismatch correction in bacteriophage lambda crosses by very short patch repair. Mol. Gen. Genet. **19**:118–125.
22. **Lloyd, R. G., F. E. Benson, and C. E. Shurvington.** 1984. Effect of ruv mutations on recombination and DNA repair in *Escherichia coli* K12. Mol. Gen. Genet. **184**:303–309.
23. **Lloyd, R. G., S. M. Picksley, and C. Prescott.** 1983. Inducible expression of a gene specific to the recF pathway for recombination in *Escherichia coli* K12. Mol. Gen. Genet. **190**:162–167.
24. **Lovett, S. T., and A. J. Clark.** 1984. Genetic analysis of the *recJ* gene of *Escherichia coli* K-12. J. Bacteriol. **157**:190–196.
25. **Lu, A. L., S. Clark, and P. Modrich.** 1983. Methyl-directed repair of DNA base mismatches *in vitro*. Proc. Natl. Acad. Sci. USA **80**:4639-4643.
26. **Mahajan, S. K., and A. R. Datta.** 1979. Mechanisms of recombination by the RecBC and RecF pathways following conjugation in *Escherichia coli* K12. Mol. Gen. Genet. **169**:67–78.
27. **Meselson, M. S., and C. M. Radding.** 1975. A general model for recombination. Proc. Natl. Acad. Sci. USA **72**:358–361.
28. **Modrich, P.** 1987. DNA mismatch correction. Annu. Rev. Biochem. **56**:435–466.
29. **Morrison, D. A.** 1977. Transformation of *Escherichia coli*: cryogenic preservation of competent cells. J. Bacteriol. **132**: 349–351.
30. **Nakayama, H., K. Nakayama, R. Nakayama, N. Irino, Y. Nakayama, and P. Hanawalt.** 1984. Isolation and genetic characterization of a thymineless death-resistant mutant of *Escherichia coli* K12: identification of a new mutation (recQ1) that blocks the RecF recombination pathway. Mol. Gen. Genet. **195**:474–480.
31. **Nevers, P., and H. C. Spatz.** 1975. *Escherichia coli* mutants uvrD and uvrE are deficient in gene conversion of lambda-heteroduplexes. Mol. Gen. Genet. **139**:233–243.
32. **Nghiem, Y., M. Cabrera, C. C. Cupples, and J. H. Miller.** 1988. The *mutY* gene: a mutator locus in *Escherichia coli* that generates G · C-T · A transversions. Proc. Natl. Acad. Sci. USA **85**:2709–2713.
33. **Potter, H., and D. Dressler.** 1982. Molecular mechanisms in genetic recombination. Annu. Rev. Biochem. **51**:727–762.
34. **Pukkila, P. J., J. Peterson, G. Herman, P. Modrich, and M. Meselson.** 1983. Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. Genetics **104**:571–582.
35. **Radding, C. M.** 1978. Genetic recombination: strand transfer and mismatch repair. Annu. Rev. Biochem. **47**:847–880.
36. **Radman, M., and R. Wagner.** 1986. Mismatch repair in *Escherichia coli*. Annu. Rev. Genet. **20**:523–538.
37. **Radman, M., R. E. Wagner, B. W. Glickman, and M. Meselson.** 1980. DNA methylation, mismatch correction and genetic stability, p. 121–130. *In* M. Alacevic (ed.), Progress in environmental mutagenesis. Elsevier/North-Holland Publishing Co., Amsterdam.
38. **Raposa, S., and M. S. Fox.** 1987. Some features of base pair mismatch and heterology repair in *Escherichia coli*. Genetics **117**:381–390.
39. **Schaaper, R. M.** 1989. *Escherichia coli* mutator *mutD5* is defective in the *mutHLS* pathway of DNA mismatch repair. Genetics **121**:205–212.
40. **Smith, G. R.** 1988. Homologous recombination in procaryotes. Microbiol. Rev. **52**:1–28.
41. **Tait, R. C., R. L. Rodrigues, and R. W. West.** 1980. The rapid purification of T4 DNA ligase from a T4 lig lysogen. J. Biol. Chem. **255**:813–815.
42. **Wagner, R., and M. Meselson.** 1976. Repair tracts in mismatched DNA heteroduplexes. Proc. Natl. Acad. Sci. USA **73**:135–139.
43. **White, R., and M. S. Fox.** 1974. On the molecular basis of high negative interference. Proc. Natl. Acad. Sci. USA **71**:1544–1548.
44. **Wildenberg, J., and M. Meselson.** 1975. Mismatch repair in heteroduplex DNA. Proc. Natl. Acad. Sci. USA **72**:2202–2206.

An *Escherichia coli* Cell-free System That Catalyzes the Repair of Symmetrically Methylated Heteroduplex DNA

R.A. FISHEL AND R. KOLODNER

Laboratory of Molecular Genetics, Dana-Farber Cancer Institute and Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

Heteroduplex DNA containing mismatched nucleotides is produced in *Escherichia coli* during the process(es) of general genetic recombination between two genetically distinct parental DNAs (Holliday 1964) and following certain types of chemical or physical mutagenesis (Drake and Baltz 1964). The repair of heteroduplex DNA to a homoduplex configuration has been used to explain the nonreciprocal recombination of closely linked markers, termed gene conversion, and map expansion (Fincham and Holliday 1970). Holliday first introduced the concept of heteroduplex DNA into a model for genetic recombination as part of a mechanism for explaining gene conversion in fungi. The success of this model has given the suggested intermediate—the Holliday structure—a central position in current mechanisms of general genetic recombination. This subject has been recently reviewed by Fox (1978), Radding (1978), Warner and Tessman (1978), and Potter and Dressler (1982).

The formation of heteroduplex DNA was originally proposed as resulting from hybrid overlap that is a necessary part of the initiation of the Holliday recombination intermediate. The extension of the heteroduplex region along the DNA is mediated by branch migration, either occurring randomly (Meselson 1972) or driven by replication or exonuclease digestion (Meselson and Radding 1975). Branch migration has been extensively studied using figure-8 DNA (Thompson et al. 1975; Warner et al. 1979), a prototype of the Holliday structure for small circular genomes, and cruciform DNA constructed from palindromic DNA sequences (Courey and Wang 1983). The observed step rate of branch migration in vitro appears to predict the presence of a catalyst in vivo to generate long enough regions of heteroduplex DNA to explain the results of genetic experiments (R.A. Fishel and R.C. Warner, unpubl.). The repair of mismatched nucleotides in heteroduplex DNA generated by the process of branch migration is an important function of genetic recombination that results in altered DNA sequence combinations.

In *E. coli*, there are at least three pathways that recognize and repair heteroduplex DNA (Fishel and Kolodner 1983; R.A. Fishel et al., unpubl.): one DNA methylation-dependent pathway, i.e., the designated *dam*-instructed repair system (Radman et al. 1980), and two DNA methylation-independent pathways. An overlap between the *dam*-instructed repair pathway and one of the DNA methylation-independent repair pathways in both protein requirement and repair mechanism appears to exist to insure an efficient response to mismatched DNA (R.A. Fishel et al., unpubl.). A distinction can be made between the two methylation-independent repair pathways on the basis of their excision tract length. The most efficient methylation-independent repair pathway is characterized by excision tracts of several thousand nucleotides and results in co-repair of most heteroallelic genetic markers. This pathway has been shown to require the *mutS* and, to a lesser extent, the *uvrD* gene products. A far less efficient pathway appears to produce excision tracts of less than 300 nucleotides and requires the *recF* and *recJ* gene products. Heteroduplex DNA formed by genetic recombination and chemical or physical mutagenesis is presumed to be symmetrically methylated and recognized by one or both of the methylation-independent mismatch repair systems.

We report here a description of a cell-free system that will catalyze the repair of symmetrically methylated heteroduplex DNA that has similar properties to those observed in vivo. In the course of our investigations into the protein requirements for heteroduplex DNA repair, we observed the induction of a potent nuclease activity in strains containing single *recJ* mutant alleles (Lovett and Clark 1984). Preliminary experiments designed to determine the identity and characteristics of this *recJ*-induced nuclease are included in this report.

METHODS

Enzymes and cofactors. Restriction endonucleases were obtained from New England Biolabs. [^{3}H]-Thymine was obtained from New England Nuclear. Unlabeled nucleotides were from P-L Biochemicals.

Strains and plasmids. All strains used in this study are derivatives of *E. coli* K12 AB1157 and are shown in Table 1. The plasmids used in this study (Table 1) are derived from pBR322 and have been described elsewhere (Doherty et al., 1983). Plasmid DNA was prepared according to James et al. (1983) and monitored for the extent of *dam* methylation using the *Dpn*I, *Mbo*I, and *Sau*3A restriction endonucleases.

Construction of heteroduplex DNA. Heteroduplex plasmid DNA was constructed by a procedure that will

Table 1. Strains and Plasmids

	Markers	Source
Strains		
AB1157	*thr1, leu6, thi1, lacy1, galk2, ara14, xyl5, mtl1, proA2, his4, argE3, str31, tsx33, supE44*	this laboratory
JC 10287	AB1157, Δ(*recA-srl*)*301*	A.J. Clark[a]
JC 9239	AB1157, *recF143*	A.J. Clark
JC 13030	AB1157, *recJ77*	A.J. Clark
JC 13031	AB1157, *recJ153*	A.J. Clark
JC 12123	AB1157, *recJ284*::Tn*10*	A.J. Clark
ES 1574	AB1157, *mutS201*::Tn*5*	E. Siegal[b]
ES 1638	AB1157, *uvrD260*::Tn*5*	E. Siegal
JC 7623	AB1157, *recB21 recC22 sbcB23*	A.J. Clark
RK 1503	AB1157, *recF143, recJ284*::Tn*10*	this laboratory
RK 1502	AB1157, *recB21, recC22, sbcB23, recJ284*::Tn*10*	this laboratory
Plasmids		
pRDK35	tet-10 (*ClaI*$^-$/*XhoI*$^+$)	Doherty et al. (1983)
pRDK37	tet-12 (*SalI*$^-$/*XhoI*$^+$)	Doherty et al. (1983)

[a]University of California, Berkeley.
[b]Tufts University, Medford, Mass.

be described in detail elsewhere. We used pRDK35 and pRDK37 parent plasmids in heteroduplex constructions because the *XhoI* linker insertion mutations inactivate unique restriction sites and permit positive detection of both the mutant and wild-type alleles. Purified heteroduplex DNA was monitored for the presence of contaminating homoduplex DNA by double digestion with *PstI* and *XhoI*, followed by electrophoretic analysis to detect characteristic parental DNA fragments. Heteroduplex DNA preparation was considered acceptable for experimental use if no detectable parental DNA fragments were observed when 0.2 μg of substrate DNA was analyzed (>98% pure).

Preparation of* E. coli *extract. Extracts of *E. coli* were prepared by a modification of the procedure described by Scott and Kornberg (1978) and Liu et al. (1983) and will be described in detail elsewhere. For experiments described in this communication, we have used a dialyzed ammonium sulfate fraction that had a protein concentration of 40–60 mg/ml.

Mismatch repair assay. Detection of heteroduplex DNA repair was monitored as the conversion of restriction-resistant DNA to restriction-sensitive DNA by gel electrophoresis. A typical reaction mix contained 10 μg/ml heteroduplex plasmid DNA, 25 mM HEPES (pH 7.8), 10 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM DTT, 1.5 mM rATP, 500 μM NAD, 100 μM (each) dNTPs, 100 μM (each) rNTPs, 100 μg/ml bovine serum albumin (BSA) in a 10-μl final reaction volume. The reaction was stopped by addition of 2 μl of 0.1 M EDTA (pH 8.0), 1.5 μl of 5 M potassium acetate, and 1.5 μl of 10% SDS to remove and precipitate the protein, followed by phenol extraction and ethanol precipitation of the DNA.

To assay for independent repair events, incubated DNA was transformed into JC9239 *recF143* according to a modification of the method originally described by Morrison (1977). Transformants were plated on Luria broth containing 1.5% agar plus ampicillin (50 μl/ml) or ampicillin + tetracycline (15 μg/ml) to determine the proportion of cells that had been transformed with a DNA molecule that had undergone a repair event to produce a wild-type tetracycline-resistant (Tc^r) gene. This method does not detect independent repair events that produce a double mutant plasmid. However, previous experiments have shown that wild-type and double mutant are produced at approximately equal frequencies (Fishel and Kolodner 1983).

Nuclease Assay. Nuclease activity was measured as an increase in trichloracetic acid (TCA)-soluble material released from circular supercoiled [^{3}H]-labeled pBR322 DNA. TCA-soluble material was assayed according to standard methods. The nuclease assay was carried out under the same conditions as those described for the heteroduplex DNA repair reaction.

RESULTS

Heteroduplex Plasmid DNA Repair in Vivo

Heteroduplex DNA introduced into *E. coli* by transformation has three possible fates, as diagramed in Figure 1: (1) DNA replication producing a mixture of both parental homoduplexes, (2) corepair characterized by an excision-resynthesis tract that covers both marker sites to produce *one* of the parental homoduplex molecules, and (3) independent repair of a single heteroduplex site, resulting in a wild-type or double mutant strand associated with one of the parental strands. Simple genetic and physical assays for the different types of heteroduplex plasmid DNA repair have been described (Fishel and Kolodner 1983). The results of a comprehensive study of the methylation and genetic requirements for the repair of heteroduplex plasmid DNA will be published elsewhere. Table 2 summarizes the results of experiments designed to determine the frequency of plasmid recombination and

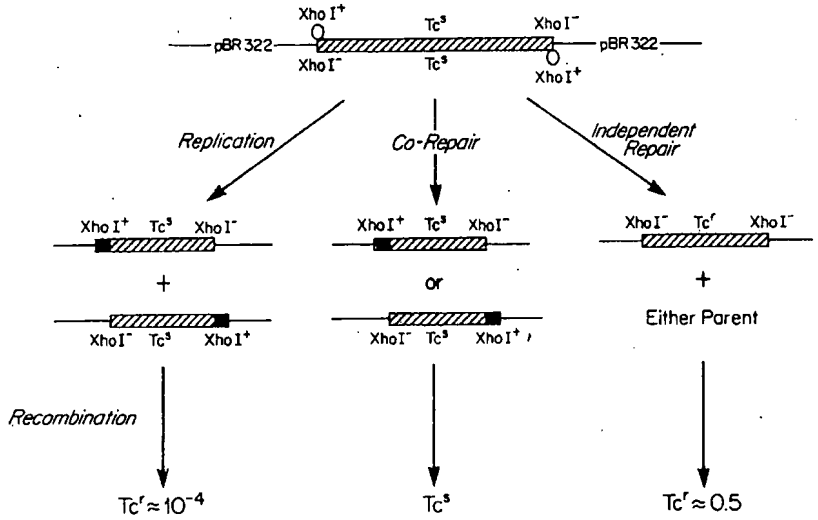


Figure 1. Illustration of the fate of heteroduplex plasmid DNA following transformation into *E. coli*.

repair of symmetrically methylated heteroduplex plasmid DNA in vivo (Fishel et al. 1981, and unpubl.; Fishel and Kolodner 1983).

Plasmid recombination requires the genes that are generally recognized as belonging to the RecF pathway shown in Table 2. These include *recA*, *recF*, *recJ*, *ssb*, and *topA* genetic loci. Plasmid recombination was unaffected by the *mutH*, *mutL*, *mutS*, and *uvrD* mutants collectively belonging to the *dam*-instructed repair pathway. An increase in the frequency of plasmid recombination in a *uvrD* strain was observed. This effect has been observed in other recombination systems and attributed to an increase in unrepaired lesions that subsequently become recombinogenic (Lloyd 1983). Independent repair of symmetrically methylated heteroduplex DNA required the *recF*, *recJ*, and *ssb* gene products, whereas co-repair required the *mutS* and *uvrD* gene products. It has been tempting to ascribe the recombination-deficient phenotype of *recF*, *recJ*, and *ssb* to a deficiency in heteroduplex DNA repair. However, the results of product analysis described by Doherty et al. (1983) indicated that plasmid recombination involves symmetric heteroduplex formation followed by co-repair or replication. These experiments tend to preclude an effect by a repair pathway that leads predominantly to independent site repair. A possible explanation of the effect of *recF*, *recJ*, and *ssb* will be described later.

Our experiments (data not shown) on the repair of hemimethylated heteroduplex plasmid DNA in vivo have confirmed the work of other laboratories (Radman et al. 1980). The methylated strand was found to be preferentially used as a template by the *dam*-instructed repair pathway, and repair required the *mutH*, *mutL*, *mutS*, and *uvrD* gene products (Liu et al. 1983; R.A. Fishel et al., unpubl.).

Table 2. The Effect of Several Mutant Strains of *E. coli* on Plasmid Recombination and Heteroduplex DNA Repair in Vivo

		Percent plasmid recombination[a]	Percent heteroduplex DNA repair[b]	
			independent	corepair
AB1157	wild type	0.1	1.16 ± 0.05	63 ± 3
	ΔrecA301	0.0006	0.85 ± 0.1	79 ± 5
	recF143	0.001	0.02 ± 0.01	75 ± 5
	recJ153	0.0008	0.01 ± 0.01	70 ± 5
	ssb113	0.005	0.26 ± 0.05	N.D.[c]
	ΔtopA700	0.001	N.D.[c]	N.D.[c]
	mutH24	0.1	0.7 ± 0.2	63 ± 3
	mutL25	0.25	0.7 ± 0.1	63 ± 8
	mutS201	0.1	0.7 ± 0.1	22 ± 7
	uvrD260	2.1	0.3 ± 0.2	37 ± 8

[a]Frequency of recombination was determined according to Fishel et al. (1981) using pRDK41 heterodimer (Doherty et al. 1983).

[b]Determined as described by Fishel and Kolodner (1983) and using symmetrically methylated heteroduplex DNA.

[c]N.D., Not determined.

Heteroduplex Plasmid DNA Repair In Vitro

A cell-free system from *E. coli* has been developed that catalyzes the repair of symmetrically methylated heteroduplex plasmid DNA. Two principal methods have been developed to measure the repair of purified heteroduplex plasmid DNA in vitro: (1) monitoring the conversion of restriction-resistant heteroduplex DNA to restriction-sensitive homoduplex DNA by gel electrophoresis and (2) quantitation of the independent repair reaction by measuring the production of plasmid DNA molecules that give rise to Tc^r colonies following transformation of incubated heteroduplex DNA into a *recF* acceptor strain. A combination of several methods is the most definitive measure of heteroduplex DNA repair.

The repair of heteroduplex DNA to homoduplex DNA produces four characteristic parental fragments following digestion of the product DNA with *Xho*I and *Pst*I (3.58 kb and 0.78 kb from the pRDK35 parent; 2.95 kb and 1.41 kb from the pRDK37 parent). Digestion with *Pst*I provided a control to demonstrate that the plasmid DNA could be digested with a restriction endonuclease after incubation in the cell-free system. Figure 2 illustrates the reaction requirements for heteroduplex DNA repair in vitro using the gel electrophoresis assay system. The reaction required rATP and dNTPs, in addition to Mg^{++} (data not shown). The repair of heteroduplex DNA did not require rGTP, rCTP, rUTP, NAD, spermidine, or Ca^{++}. We have used dialyzed ammonium sulfate fractions in these experiments; however, these protein fractions may contain some residual nucleotides or cofactors that could substitute for the added cofactors. Thus, the reaction requirements shown in Figure 2 should be regarded as preliminary until a more purified system is examined. Depending on the amount and type of extract used, we have found that 30–50% of the heteroduplex substrate is repaired in vitro. It appears that our extracts are less efficient at the repair of symmetrically methylated heteroduplex DNA than the methyl-directed repair system reported by Liu et al. (1983) and may reflect the background of methyl-independent repair observed by these workers.

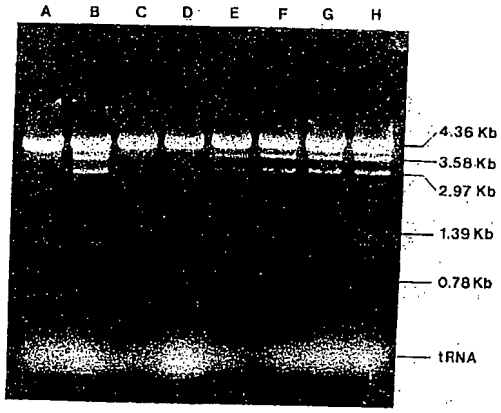


Figure 2. Reaction requirements for heteroduplex DNA repair in vitro. Incubation performed as described in methods section. (Lane *A*) Complete reaction mix, no incubation; (lane *B*) complete reaction mix; (lane *C*) minus rATP; (lane *D*) minus the dNTPs; (lane *E*) minus rGTP, rCTP, rUTP; (lane *F*) minus NAD; (lane *G*) minus spermidine; (lane *H*) minus calcium chloride.

Incubation of heteroduplex DNA in the cell-free system followed by transformation into AB1157 *recF143* increased the number of Tc^r transformants obtained by approximately 10-fold (Table 3). Reactions with extracts prepared from a strain containing the *recF143* mutation yielded a lower proportion of Tc^r transformants and is consistent with the results of experiments performed in vivo. There was no effect on the number of Tc^r transformants when the heteroduplex DNA substrate was incubated with an extract prepared from *E. coli* cells containing a *mutS* mutation. These results imply that there is a component of heteroduplex DNA repair in vitro that leads to independent repair of mismatch sites and is dependent on a functional *recF* gene product. We have verified the independent repair of mismatch sites by excising the 3.58-kb fragment produced in vitro and digesting it with *Sal*I. This analysis demonstrated that a fraction of DNA molecules was resistant to digestion by *Sal*I (data not shown). Sensitivity to *Sal*I indicates that the (*Sal*I$^-$/*Xho*I$^+$) heteroduplex site has been corepaired to homoduplex during the repair to *Xho*I$^+$ of the pRDK35 (*Xho*I$^+$/*Cla*I$^-$) heteroduplex site whereas resistance to digestion by both *Xho*I and *Sal*I indicates that the pRDK37 (*Sal*I$^-$/*Xho*I$^+$) site was not repaired (note that *Xho*I sensitivity is used to isolate the 3.58-kb fragment). These results are consistent with both independent repair and co-repair reactions occurring in vitro.

The observation that the *mutS* mutation had no effect on independent repair events in vitro (Table 3) is consistent with results obtained in vivo. However, the *mutS* mutation was shown to reduce repair when assayed by gel electrophoresis (Fig. 3) suggesting a role in the co-repair reaction in vitro.

The repair of symmetrically methylated heteroduplex DNA in vitro was shown to be independent of the *recA* and *uvrD* gene products (Fig. 3). The independence of mismatch nucleotide repair on *recA* function is consistent with observations in vivo. The independence of mismatch nucleotide repair in vitro on *uvrD* may reflect the fact that there are several helicase activities in *E. coli* that might substitute for *uvrD* (helicase II) activity (Maples and Kushner 1982). Hetero-

Table 3. Frequency of Tetracycline Resistance Generated in Vitro

Extract	Percent Tet$^+$[a]
Unincubated	0.019
Wild type	0.12
recF143	0.03
mutS201::Tn*5*	0.19

[a]Percent Tet$^+$, number of Tet$^+$ colonies/total number of transformants, following transformation of the product DNA into JC9239 *recF143* strain.

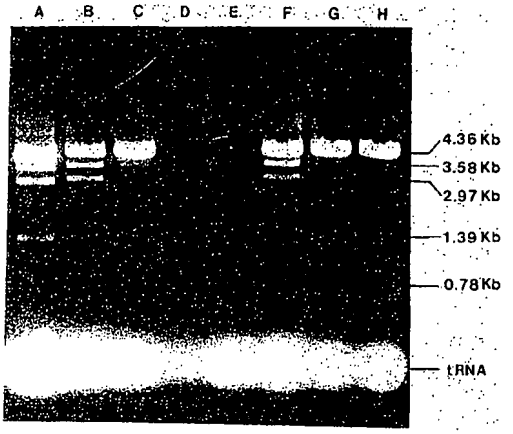


Figure 3. Genetic requirements for heteroduplex DNA repair in vitro. Reactions were performed according to described methods. (Lane *A*) 335 μg of wild-type extract; (lane *B*) 295 μg of *recA301* extract; (lane *C*) 425 μg of *recF143* extract; (lane *D*) 255 μg of *recJ77* extract; (lane *E*) 325 μg of *recJ153* extract; (lane *F*) 342 μg of *uvrD* 260 extract; (lane *G*) 338 μg of *mutS201* extract; (lane *H*) 213 μg of *recF143* extract and 169 μg of *mutS201* extract.

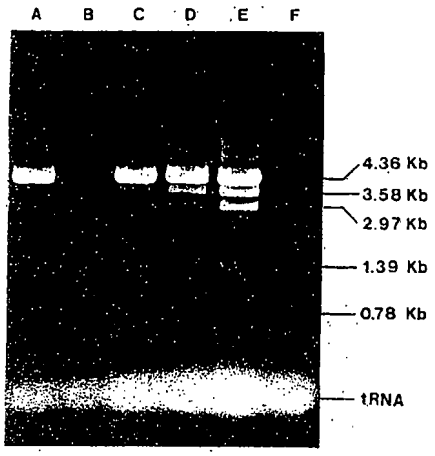


Figure 4. Genetic requirements for the *recJ*-induced nuclease. Repair reactions in vitro were performed according to the described methods and using 300 μg of total protein. (Lane *A*) *recF143*; (lane *B*) *recJ153*; (lane *C*) *recF143 recJ284; (lane D) recB21 recC22 sbcB23; (lane E) recB21 recC22 sbcB23 recJ284*; (lane *f*) *recF143* + *recJ153*.

duplex DNA repair in vitro was shown to be dependent on the *recF* and *mutS* genes, confirming results obtained in vivo. When *recF* and *mutS* extracts were mixed, the ability of these extracts to carry out the repair reaction in vitro was not restored. Repair activity is present in vitro when either *mutS* or *recF* extracts are mixed with a wild-type extract (data not shown), eliminating the possibility that the *recF* or *mutS* extracts contain inhibitors of repair activity. These results suggest that the effect of *mutS* and *recF* mutations in vitro may be more complex than can be explained by the simple absence of only one protein in each mutant strain. The two *recJ* strains tested resulted in complete loss (apparent degradation) of the heteroduplex DNA (Fig. 3). This result suggested that there might be a potent nuclease present in *recJ* extracts.

recJ-Induced Nuclease

Strains containing a mutation in the *recF* or *recJ* genes have been shown to be deficient in conjugation-mediated recombination in a *recB recC sbcB* background and deficient in plasmid recombination and the independent repair of heteroduplex plasmid DNA when present as single mutations in an otherwise wild-type background. Unlike the *recF* single mutation, the *recJ* single mutation has no known phenotype (Lovett and Clark 1984). On the basis of experiments shown in Figure 3, we investigated heteroduplex DNA repair activity and the apparent loss of DNA in several *recJ* strain derivatives. Figure 4 shows that the extracts prepared from *recF recJ* and *recB recC sbcB recJ* strains did not degrade heteroduplex plasmid DNA. In addition, the *recB recC sbcB recJ* strain was shown to be proficient in the repair reaction, suggesting a complex relationship between DNA repair and nuclease activity. The mixing of extracts from a *recF* (see Fig. 4) or *recB recC sbcB* strain (data not shown) with the *recJ* extract had no effect on DNA degradation, suggesting that the effect of these mutations is not directly inhibitory.

The loss of DNA does not necessarily imply degradation. To address directly the question of degradation, we determined that *recJ* extracts will digest any circular duplex DNA to TCA-soluble material. Figure 5 illustrates the conversion of circular supercoiled [^{3}H]pBR322 to TCA-material following incubation with extracts prepared from wild-type and mutant strains of *E. coli*. The specific activity of the *recJ*-induced "nuclease" is shown to increase approximately 10-fold when a constant amount of a *recF recJ* extract is included in the reaction. This result implies that the *recF recJ* extract can supply a portion of the "nuclease" activity or a stimulating factor and suggests that degradation is a multistep process.

The *recJ*-induced "nuclease" activity was found to be decreased substantially if rATP was left out of the reaction mix (data not shown). The dependence of the *recJ*-induced nuclease activity on rATP suggests the direct participation of exonuclease V (*recB recC* gene product), since it is the only known rATP-dependent exonuclease activity in *E. coli* (Goldmark and Linn 1972). Alternatively, the *recJ* mutation might induce a novel rATP-independent endonuclease or exonuclease or the reaction might require some other rATP-dependent protein. A model for DNA degradation involving an endonuclease or gapping activity not present in *recF recJ* extracts and exonuclease V activity, which is not present in *recB recC sbcB recJ*, is not plausible since a *recF recJ* extract was unable to complement a *recB recC sbcB recJ* extract (see Fig. 5). This

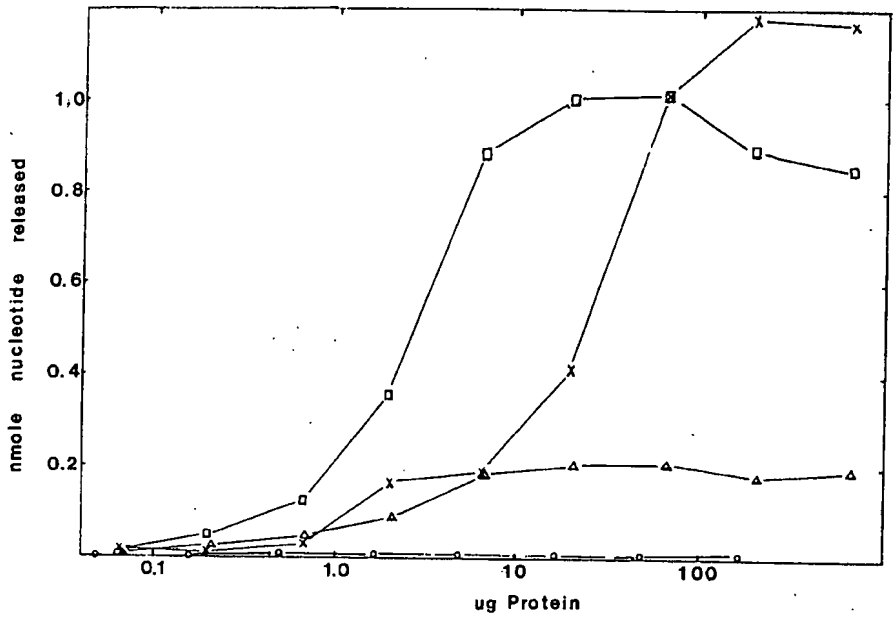


Figure 5. Nuclease activity on covalently closed circular, supercoiled plasmid DNA. (△) Wild-type extract; (×) *recJ153* extract; (□) *recJ153* extract + 35 μg of *recF143 recJ284* extract; (○) *recB21 recC22 sbcB23 recJ284* extract + 35 μg of *recF143 recJ284* extract.

experiment suggests that the induced degradation of circular DNA by *recJ* mutations may be the result of a complex regulatory process.

DISCUSSION

We have presented evidence that symmetrically methylated heteroduplex DNA is efficiently repaired to homoduplex DNA both in vivo and in vitro. The process(es) of repair are independent of *recA* function, consistent with a role for RecA protein that is prior to heteroduplex repair (Radding 1978). The most efficient pathway for repair results in the co-repair of heteroduplex sites that are separated by up to 1243 bp, suggesting excision-resynthesis tracts that cover several thousand nucleotides (Fishel and Kolodner 1983). The repair of symmetrically methylated heteroduplex DNA requires the *mutS* gene product originally identified as a component of the *dam*-instructed repair pathway. Overlap of the methylation-independent pathway of repair with the *dam*-instructed pathway (see Pang et al.; Lu et al.; both this volume) insures repair of heteroduplex DNA even if a methylation asymmetry does not exist near the mismatch. The *dam*-instructed pathway has been shown to be more efficient than the methylation-independent pathway, which is likely to promote increased replication fidelity, unless there is no methylation asymmetry, in which case repair may be mutagenic.

Independent repair was shown to require the *recF* gene product both in vivo and in vitro. In addition, *recF* has been shown to reduce plasmid recombination in vivo. The results of Doherty et al. (1983) suggest that the function of *recF* is more complex than a singular role in the resolution of heteroduplex DNA because most plasmid recombination events involved co-repair rather than independent repair. One possible rectification of the role of *recF* would suggest that the independent repair of mismatched nucleotides is a secondary activity. A more general role for *recF* in the RecF pathway would be to create a recombinogenic structure from circular, supercoiled, duplex DNA such as a gap or displaced DNA strand. Such an activity could be involved in repair of a mismatch site and thus account for a low frequency of independent repair events. Since our experiments have shown that co-repair of mismatched nucleotides is the most frequent repair event, wild-type recombinants observed following heteroallelic crosses are probably the result of single-marker heteroduplex followed by co-repair or replication as proposed by Doherty et al. (1983).

We have identified a "nuclease" activity present in *recJ* single mutant strains that digests circular duplex DNA. The rATP-dependent exonuclease V appears to be at least partially responsible for this activity, which is absent in both *recF recJ* and *recB recC sbcB recJ* strains. The *recJ* mutation has been shown to be involved in the RecF pathway during genetic recombination and in independent repair of heteroduplex DNA. It is possible that the degradation of circular duplex DNA, reduced genetic recombination, and reduced heteroduplex repair activity in *recJ* strains are related. An extension of the role of *recJ* discussed above proposes that the *recJ* protein protects or plays a positive regulatory role in the protection of intermediates, produced during mismatch repair or during *recF*-mediated recombination, from degradation by

exonuclease V or other enzymes. It is, however, interesting to note that the *recJ*-induced "nuclease" activity cannot apparently be reconstituted from two extracts that have lost this activity as a result of the introduction of different secondary mutations. These experiments point to the complexity of recombination and repair processes and define the need for further biochemical and genetic analysis of the proteins involved.

ACKNOWLEDGMENTS

The authors would like to thank A.J. Clark for many helpful discussions. This work was supported by NIH postdoctoral fellowship GM07693 to R.A.F. and NIH grant GM26017 and ACS grants JFRA-35 and FRA-271 to R.K.

REFERENCES

COUREY, A.J. and J.C. WANG. 1983. Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. *Cell* **33:** 817.

DOHERTY, M.J., P.T. MORRISON, and R. KOLODNER. 1983. Genetic recombination of bacterial plasmid DNA: Physical and genetic analysis of the products of plasmid recombination in *Escherichia coli*. *J. Mol. Biol.* **167:** 539.

DRAKE, J.W. and R.H. BALTZ. 1976. The biochemistry of mutagenesis. *Annu. Rev. Biochem.* **45:** 11.

FINCHAM, J.R.S. and R. HOLLIDAY. 1970. An explanation of fine structure map expansion in terms of excision repair. *Mol. Gen. Genet.* **109:** 309.

FISHEL, R.A. and R. KOLODNER. 1983. Gene conversion in *Escherichia coli*: The identification of two repair pathways for mismatched nucleotides. *UCLA Symp. Mol. Cell. Biol.* **11:** 309.

FISHEL, R.A., A.A. JAMES, and R. KOLODNER. 1981. *recA*-independent general genetic recombination of plasmids. *Nature* **294:** 184.

FOX, M.S. 1978. Some features of genetic recombination in procaryotes. *Annu. Rev. Genet.* **12:** 47.

GOLDMARK, P.J. and S. LINN. 1972. Purification and properties of the *rec*BC DNase of *Escherichia coli* K12. *J. Biol. Chem.* **247:** 1849.

HOLLIDAY, R. 1964. A mechanism for gene conversion in fungi. *Genet. Res.* **5:** 282.

JAMES, A.A., P.T. MORRISON, and R. KOLODNER. 1983. Genetic recombination of bacterial plasmid DNA: Analysis of the effect of recombination-deficient mutations on plasmid recombination. *J. Mol. Biol.* **160:** 411.

LIU, A.-L., S. CLARK, and P. MODRICH. 1983. Methyl-directed repair of DNA base-pair mismatches *in vitro*. *Proc. Natl. Acad. Sci.* **80:** 4639.

LLOYD, R.G. 1983. *lex*A dependent recombination in *uvr*D strains of *Escherichia coli*. *Mol. Gen. Genet.* **189:** 157.

LOVETT, S.T. and A.J. CLARK. 1984. Genetic analysis of the *rec*J gene of *Escherichia coli* K12. *J. Bacteriol.* **157:** 190.

MAPLES, V.F. and S.R. KUSHNER. 1982. DNA repair in *Escherichia coli*: Identification of the *uvr*D gene product. *Proc. Natl. Acad. Sci.* **79:** 5616.

MESELSON, M. 1972. Formation of hybrid DNA by rotary diffusion during genetic recombination. *J. Mol. Biol.* **71:** 795.

MESELSON, M.S. and C.M. RADDING. 1975. A general model for genetic recombination. *Proc. Natl. Acad. Sci.* **72:** 358.

MORRISON, D.A. 1977. Transformation in *Escherichia coli*: Cryogenic preservation of component cells. *J. Bacteriol.* **132:** 349.

POTTER, H. and D. DRESSLER. 1982. Molecular mechanisms in genetic recombination. *Annu. Rev. Biochem.* **51:** 727.

RADDING, C.M. 1978. Genetic recombination: Strand transfer and mismatch repair. *Annu. Rev. Biochem.* **47:** 847.

RADMAN, M., R.E. WAGNER, Jr., B.W. GLICKMAN, and M. MESELSON. 1980. DNA methylation, mismatch correction and genetic stability. In *Progress in environmental mutagenesis* (ed. M. Alacevic), vol. 7, p. 121. Elsevier/North-Holland, Amsterdam.

SCOTT, J.F. and A. KORNBERG. 1978. Purification of the *rep* protein of *Escherichia coli*. *J. Biol. Chem.* **253:** 3292.

THOMPSON, B.J., C. ESCARMIS, B. PARKER, W.C. SLATER, J. DONIGER, I. TESSMAN, and R.C. WARNER. 1975. Figure-8 configuration of dimers of S13 and ϕX174 replicative form DNA. *J. Mol. Biol.* **91:** 409.

WARNER, R.C. and I. TESSMAN. 1978. Mechanisms of genetic recombination. In *The single-strand DNA phages* (ed. D.T. Denhardt et al.), p. 417. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

WARNER, R.C., R.A. FISHEL, and F.C. WHEELER. 1979. Branch migration in recombination. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 957.

MOLECULAR AND CELLULAR BIOLOGY, Jan. 1987, p. 218–224
0270-7306/87/010218-07$02.00/0

Vol. 7, No. 1

DNA Mismatch Repair Detected in Human Cell Extracts

PETER M. GLAZER, SAUMYEN N. SARKAR, GEORGE E. CHISHOLM, AND WILLIAM C. SUMMERS*

Departments of Therapeutic Radiology, Human Genetics, and Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06510

Received 31 March 1986/Accepted 19 September 1986

A system to study mismatch repair in vitro in HeLa cell extracts was developed. Preformed heteroduplex plasmid DNA containing two single base pair mismatches within the *SupF* gene of *Escherichia coli* was used as a substrate in a mismatch repair assay. Repair of one or both of the mismatches to the wild-type sequence was measured by transformation of a *lac*(Am) *E. coli* strain in which the presence of an active *supF* gene could be scored. The *E. coli* strain used was constructed to carry mutations in genes associated with mismatch repair and recombination (*mutH*, *mutU*, and *recA*) so that the processing of the heteroduplex DNA by the bacterium was minimal. Extract reactions were carried out by the incubation of the heteroduplex plasmid DNA in the HeLa cell extracts to which ATP, creatine phosphate, creatine kinase, deoxynucleotides, and a magnesium-containing buffer were added. Under these conditions about 1% of the mismatches were repaired. In the absence of added energy sources or deoxynucleotides, the activity in the extracts was significantly reduced. The addition of either aphidicolin or dideoxynucleotides reduced the mismatch repair activity, but only aphidicolin was effective in blocking DNA polymerization in the extracts. It is concluded that mismatch repair in these extracts is an energy-requiring process that is dependent on an adequate deoxynucleotide concentration. The results also indicate that the process is associated with some type of DNA polymerization, but the different effects of aphidicolin and dideoxynucleotides suggest that the mismatch repair activity in the extracts cannot simply be accounted for by random nick-translation activity alone.

The repair of base pair mismatches in the DNA of an organism plays an important role in reducing the frequency of mutations and in preserving the genetic integrity of the organism. Mismatches can occur in several ways. Recombination events can generate heteroduplex regions in DNA, and the process of gene conversion is thought to involve heteroduplex structures as intermediates. In DNA replication errors can occur which produce mismatched bases that must be corrected to avoid a high rate of spontaneous mutagenesis.

In *Escherichia coli* DNA mismatch repair has been studied extensively. Transfection experiments with heteroduplex bacteriophage DNA (for a review, see reference 11) have demonstrated that *E. coli* has an efficient system for mismatch repair. The power of procaryotic genetic analysis has allowed identification of mutants that are deficient in various aspects of the repair process, and this has led to an understanding of some of the mechanisms that are involved. The products of the *mutH*, *mutL*, *mutS*, and *mutU* loci all seem to play a role in mismatch repair. Mutations at these loci produce strains that undergo a high rate of spontaneous mutagenesis because they cannot repair DNA mismatches effectively (11). Experiments involving the transfection of *E. coli* with hemimethylated λ DNA coupled with the identification of mutants with altered function in the methylase encoded by the *dam* gene have led to a model of methyl-directed strand selection in *E. coli* mismatch repair (5, 8, 14, 17). The subsequent development of a system to detect mismatch repair in cell extracts of *E. coli* has greatly facilitated the study of the enzymology of this process (12).

In contrast, much less is known about the mechanisms of heteroduplex repair in mammalian cells. Microinjection experiments have demonstrated that mouse cells can efficiently correct mismatched bases in exogenously prepared heteroduplex DNA (4). Studies involving the transfection of monkey cells with hemimethylated simian virus 40 (SV40) DNA have suggested that strand selection in mismatch repair in these cells may be influenced by methylation patterns (7).

Extension of these findings by the sort of genetic analysis used with *E. coli*, however, is hampered by the difficulty of genetic manipulation of mammalian cells. As another approach, we sought to develop an assay to detect DNA mismatch repair in cell extracts of mammalian cells as a way to study the mechanisms of repair directly in vitro. Such an approach to the study of other aspects of mammalian DNA and RNA metabolism has already been described in several studies (2, 9, 10, 13, 23). We report here our work with extracts from HeLa cells in which we were able to detect repair of DNA mismatches in specifically constructed substrate heteroduplex DNA. Results of preliminary experiments indicate that this activity is dependent on ATP, deoxynucleotides, and DNA polymerization.

MATERIALS AND METHODS

Cells. HeLa cells were obtained from P. Ghosh, Yale University (New Haven, Conn.). The construction of *E. coli* SY204 *lacZ125*(Am) *trp-49*, *hsdR2*::Tn*10* has been described previously (20). *E. coli* EG826 *lacZ125*(Am) *trp-49 hsdR2*::Tn*10 ssb-1 malE*::Tn*10* was made by Efim Golub by P1 transduction of the *ssb-1* mutation (15) into SY204. SY208 *lacZ125*(Am) *hsdR2*::Tn*10 mutH3 mutU4 strA143* was constructed from E. coli KL874 F⁻ *hisF818 leu3 lacZ498*, *rpsL143 mutH3 mutU4* by first introducing a deletion spanning the *lac* and *pro* loci by conjugation and then introducing the *lacZ125*(Am) mutation by a second conjugation, followed by introduction of the host restriction mutation *hsdR2* by P1 transduction with Tn*10* tetracycline resistance. SY209 was constructed by P1 transduction of the *recA56* mutation from *E. coli* MC16 *argH trpA36 srl-300*::Tn*10 recA56* into SY208 by contransduction of Tn*10* as a screen after first

* Corresponding author.

curing SY208 of tetracycline resistance. SY302 was constructed by P1 transduction of the *recA56* mutation, again from *E. coli* MC136, into SY204 by cotransduction of Tn*10* as a screen after first curing SY204 of tetracycline resistance.

Plasmids. Plasmid p3AC was constructed as described previously (20). Plasmids p3AC-4 and p3AC-8 differ from p3AC only in that they bear single point mutations in the amber suppressor tyrosine tRNA gene of *E. coli*, *supF* (1), rendering this gene nonfunctional. Plasmids p3AC-4 and p3AC-8 were isolated in the course of a study of mutagenesis described previously (20). The *supF* genes in these plasmids were sequenced directly from the plasmid DNA by the method of Sanger et al. (19) with a pBR322 *Eco*RI site primer (New England Biolabs, Inc., Beverly, Mass.).

Heteroduplex preparation. For the preparation of heteroduplex molecules from p3AC-4 and p3AC-8, 25 μg of p3AC-4 linearized at the *Sca*I site and p3AC-8 linearized at the *Bam*HI site were mixed in a total volume of 1 ml of water to which 110 μl of 1 N NaOH was added. After 30 min at room temperature, 110 μl of 1 M NaH_2PO_4 and 1,280 μl of deionized formamide were added. The solution was incubated overnight at 37°C, followed by overnight dialysis against 10 mM Tris and 1 mM EDTA (pH 8). The DNA in the sample was concentrated by ethanol precipitation and examined by agarose gel electrophoresis for successful generation of the nicked circular duplexes representing heteroduplex molecules, prior to use in the experiments described below.

Transformations. Bacterial transformations were done by the method of Hanahan (6). Bacteria transformed to ampicillin resistance were screened for β-galactosidase activity by growth in the presence of the chromogenic indicator, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) at 50 μg/ml and isopropyl β-D-thiogalactopyranoside at 20 μg/ml.

HeLa cell extracts. Suspension cultures of HeLa cells were grown in RPMI 1640 medium (GIBCO Laboratories, Grand Island, N.Y.) supplemented with 10% fetal calf serum. Whole cell extracts, from 2 to 3 liters of culture containing 4.5×10^5 to 5×10^5 cells per ml, were prepared essentially as described by Manley et al. (13) with some minor modifications. Cellular material precipitated by 60% ammonium sulfate was collected and suspended in the prescribed volume of a buffer containing 50 mM Tris (pH 7.9), 6 mM $MgCl_2$, 0.2 mM EDTA, 40 mM $(NH_4)_2SO_4$, 15% glycerol, and 1 mM dithiothreitol. This solution was dialyzed for 15 h at 4°C against two 500-ml volumes of 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid; pH 7.9), 100 mM KCl, 12.5 mM $MgCl_2$, 0.1 mM EDTA, 17% glycerol, and 1 mM dithiothreitol. Precipitated material was removed by centrifugation (10 min at 12,000 × *g* at 4°C), and the resulting supernatant was quick frozen in fractions and stored at −80°C. Protein concentrations of the extracts were 10 to 12 mg/ml. These extracts were also used for in vitro transcription reactions and were found to be active in this assay, producing predicted, specific, template-dependent products.

Reaction conditions. Reactions were carried out at 37°C for 2 h in a total volume of 25 μl containing 15 μl of extract (150 to 180 μg of protein), between 100 and 500 ng of DNA in 5 μl of water, and 5 μl of the appropriate buffer. The final concentrations of components in the complete reaction were 12 mM HEPES (pH 7.9); 60 mM NaCl; 60 mM KCl; 7.5 mM $MgCl_2$; 3 mM $MgSO_4$; 0.1 mM each of dGTP, dATP, dTTP, and dCTP; 0.9 mM ATP; 10 mM creatine phosphate; 10 μg of creatine kinase per ml; 0.2 mM dithiothreitol; and 10% glycerol. Reactions were terminated by the addition of 125 μl of 10 mM Tris (pH 8)–5 mM EDTA–0.1% sodium dodecyl sulfate–200 μg of proteinase K per ml. After 1 h at 37°C, protein was extracted from the samples with phenol-chloroform, and the DNA was precipitated with ethanol, redissolved, and used for bacterial transformations.

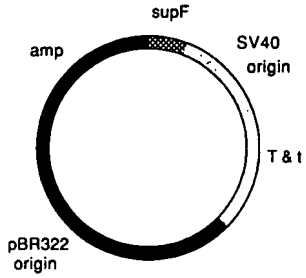


FIG. 1. Structure of the plasmid p3AC. This plasmid contains a total of 6,128 base pairs. It was constructed by ligation of a 200-base-pair *Eco*RI fragment containing the *supF* gene from the plasmid πVX into the unique *Eco*RI site of the plasmid pBR322. In addition, the 622-base-pair *Hae*II B fragment of pBR322 was deleted, and a fragment containing the *Bam*HI to *Hpa*II early region of SV40 (2,187 base pairs) was inserted at the *Cla*I site which is present in pBR322.

DNA polymerase assays. Extract reactions were carried out as described above, except that unlabeled dCTP was omitted and [α-^{32}P]dCTP was added to all reactions at a concentration of 20 μM and a specific activity of 3,000 Ci/mmol. The buffer was adjusted according to the desired experimental conditions. Just prior to the phenol extraction step, fractions of each sample either were spotted onto filter disks for measurement of radioactivity incorporated into trichloroacetic acid-insoluble material or were subjected to agarose gel electrophoresis and autoradiography.

RESULTS

Experimental design. The study of mismatch repair in mammalian cell extracts depends on a method of detecting and measuring such repair. We chose to develop a biological assay that would exploit the power of *E. coli* genetics. This entailed the generation of mutations in a gene which has a discernible phenotype in *E. coli* and the use of mutant genes to prepare heteroduplex DNA. This heteroduplex DNA was incubated in mammalian cell extracts, and the DNA was recovered from the extracts and used to transform a suitable strain of *E. coli*. Analysis of the phenotypes of the transformed *E. coli* allowed detection of heteroduplex repair. A crucial aspect of this assay was the construction of a strain of *E. coli* which was deficient in the metabolism of the heteroduplex DNA. This was needed so that the processing of the heteroduplex DNA could be attributed to reactions that occurred in the mammalian cell extracts rather than to in vivo repair in the bacterial cells.

Heteroduplex preparations. As a basis for the heteroduplex assay, we used the plasmid p3AC (Fig. 1). This plasmid contains the pBR322 origin of replication and the ampicillin resistance gene, along with the simian virus 40 (SV40) replication origin and T-antigen gene. It also contains the *supF* gene, which is an amber suppressor, tyrosine tRNA gene of *E. coli*. When a plasmid bearing a functional *supF* gene is introduced into an *E. coli* strain that has an amber nonsense mutation in the β-galactosidase gene and the resulting bacterial colonies are grown in the presence of the chromogenic β-galactosidase indicator X-gal, the colonies

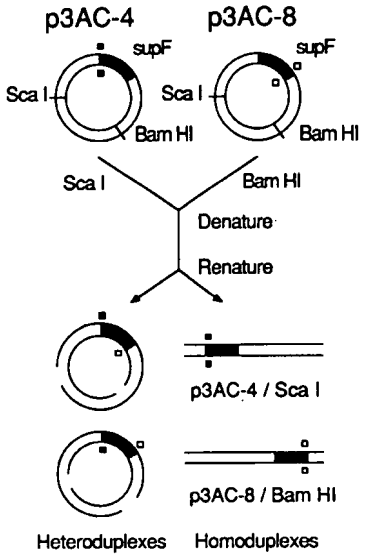


FIG. 2. Preparation of heteroduplex plasmid DNA. Plasmids p3AC-4 and p3AC-8, each bearing a single mutation in the *supF* gene as indicated by the open and closed boxes, were linearized with *Sca*I and *Bam*HI, respectively. The linearized plasmids were mixed, denatured, and renatured. In this process, circular molecules are formed which represent heteroduplexes containing two single base pair mismatches in the *supF* gene.

are blue. When the *supF* gene is absent or is nonfunctional, the colonies that are formed are white. We isolated derivatives of the plasmid p3AC in which single point mutations were introduced into the *supF* gene, eliminating its suppressor activity. The sequences of the mutant *supF* genes were determined, and two plasmids with mutant *supF* genes were chosen for this study. The mutation contained in the *supF* gene in plasmid p3AC-4 is a T to C transition, while that in p3AC-8 is a C to T transition at a site in the gene that is 61 base pairs away.

From p3AC-4 and p3AC-8 heteroduplex molecules each containing two single base pair mismatches within the *supF* gene were constructed. The scheme for construction of the heteroduplexes is shown in Fig. 2. The circular plasmids were separately converted to linear molecules by digestion with different restriction enzymes which cut each plasmid only at one site. The linear molecules were mixed, denatured, and allowed to renature. In the renaturation step, some strands from p3AC-4 annealed to strands from p3AC-8, because they were homologous except for 2 of about 6,200 base pairs. Because the plasmids were linearized at different sites, annealing between strands from the different plasmids yields linear molecules with large, complementary, single-stranded overhangs. These single-stranded regions can anneal among themselves, combining either with the complementary single strand on the same molecule or with one from another molecule, yielding nicked circular molecules or multimers, respectively. In contrast, reannealing of strands from the same plasmid regenerates the original linear molecules.

The results of the hetroduplex preparation from p3AC-4 and p3AC-8 are illustrated in Fig. 3, which shows an analysis by agarose gel electrophoresis of the steps in this process. Lanes 1 and 2 show the uncut plasmids, while the plasmids digested with their respective restriction enzymes are shown in lanes 3 and 4. As a control p3AC-8 was linearized with

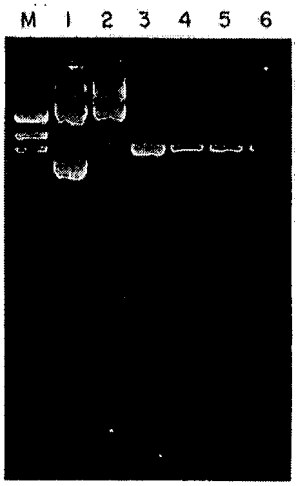


FIG. 3. Analysis by agarose gel electrophoresis of the preparation of heteroduplex plasmid DNA. M, λ-*Hin*dIII size markers; lane 1, p3AC-4; lane 2, p3AC-8; lane 3, p3AC-4–*Sca*I; lane 4, p3AC-8–*Bam*HI; lane 5, p3AC-8-*Bam*HI denatured and renatured by itself; lane 6, p3AC-4–*Sca*I and p3AC-8–*Bam*HI mixed, denatured, and renatured, as illustrated in Fig. 2.

*Bam*HI, denatured, and allowed to renature by itself in the absence of p3AC-4 (lane 5). The results of the heteroduplex preparation, in which the two linearized plasmids were denatured and renatured together, is presented in lane 6. The new band of reduced mobility (relative to the linear molecules) in lane 6 represents the nicked, circular heteroduplex molecules. The nicked, circular molecules are similar to plasmid p3AC, except for the presence of two single base pair mismatches within the *supF* gene. Because these molecules are formed by the combination of either of the two strands in p3AC-4 with its complement in p3AC-8, there are two possible heteroduplex molecules (Fig. 4), each with two mismatches, that are presumably formed in equal amounts. In either case note that both strands bear a base change that inactivates the *supF* gene. Semiconservative replication of the unrepaired heteroduplexes would yield the original mutant plasmids with defective *supF* genes. In the absence of postreplicative recombination, it is only by repair of one or both strands to the normal base prior to replication that a functional *supF* gene can be generated.

***E. coli* strains.** Our experimental goal was to use hetroduplex DNA bearing mismatches in the *supF* gene as a

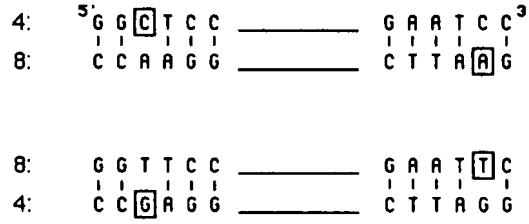


FIG. 4. Partial sequence of the heteroduplex plasmids made from p3AC-4 (designated 4) and p3AC-8 (designated 8) showing the base pair mismatches. The nucleotides indicated by the squares represent mutations from the wild-type *supF* sequence. The two mismatches within each molecule are separated by 61 base pairs. The strands in the heteroduplexes originating from either p3AC-4 or p3AC-8 are indicated.

TABLE 1. Phenotypes of colonies formed by transformation of *E. coli* mutants with heteroduplex plasmid DNA

E. coli strain	Relevant genotype	No. blue	Total	% blue
SY204	Wild type	166	8,350	2.00
EG826	*ssb-1*	17	1,970	0.86
SY302	*recA56*	24	2,661	0.90
SY208	*mutH3 mutU4*	118	5,730	2.06
SY209	*mutH3 mutU4 recA56*	37	22,235	0.17

substrate for assaying mismatch repair in mammalian cell extracts. The repair of one or both of the mismatches in the heteroduplex plasmid DNA could be detected by using the DNA, after incubation in the extracts, to transform an *E. coli* strain bearing an amber mutation in the β-galactosidase gene to ampicillin resistance in the presence of X-gal. The appearance of blue colonies would indicate a newly generated functional *supF* gene. *E. coli*, however, is able to repair mismatches in DNA. In addition, replication of unrepaired heteroduplex molecules followed by recombination between the resulting mutant plasmids could generate plasmids with functional *supF* genes. Hence, transformation of *E. coli* with heteroduplex DNA in the absence of prior incubation with cell extracts could yield a significant proportion of blue colonies because of metabolism within the bacteria. Because mismatch repair and recombination in *E. coli* are efficient processes, the effect of a mammalian cell extract on the heteroduplex DNA might be too small to detect above the background in the assay from bacterial processing of the heteroduplex.

To circumvent the background problem, we constructed several *E. coli* strains with mutations in some of the genes that are thought to play a role in mismatch repair, recombination, or both. It was necessary that all strains also contain amber mutations in the β-galactosidase gene to be useful in the assay for *supF* activity. These strains were transformed to ampicillin resistance with heteroduplex DNA, and the number of blue colonies as a percentage of the total in each case was determined. Table 1 gives the results of this experiment, along with the relevant genotypes of the strains studied. A mutation in the gene for the single-stranded binding protein of *E. coli*, as in EG826, has a measurable effect on the metabolism of heteroduplex plasmids in *E. coli*. A similar effect was seen as a consequence of a mutation in the *recA* gene, as in SY302. In contrast, the presence of mutations in both the *mutH* and *mutU* genes, as in SY208, has no detectable effect on the outcome of this assay relative to the outcome in the wild type. The presence of an additional mutation in the *recA* gene along with mutations in the *mutH* and *mutU* genes, as in SY209, however, reduces the percentage of blue colonies that are produced by a factor of 12 relative to both the wild type and the *mutH mutU* double mutant and by a factor of about 5 relative to the *recA*-deficient strain.

These results suggest that the *recA* protein has a significant role in the processing of heteroduplex plasmid DNA, probably in postreplicative recombination, but perhaps also in the mismatch repair process itself. The *mutH* and *mutU* gene products also play a role in the processing of the heteroduplex plasmid DNA, leading to the generation of functional *supF* genes, but in this assay their role is manifest only in the presence of a mutation in the *recA* gene. It should be noted that this heteroduplex DNA is derived from plasmids grown in *E. coli* SY204, which is a wild type with respect to the *E. coli* DNA methylation systems. Hence, the heteroduplex DNA is fully methylated (according to the *E. coli* pattern) on both strands. Because the mismatch repair system in *E. coli* that is related to the *mutH* and *mutU* genes is normally guided by differences in the methylation patterns on the strands of a heteroduplex, as in newly synthesized, hemimethylated DNA, the limited effect of the *mutH* and *mutU* mutations in this assay might be related to the fact that the heteroduplex DNA is fully methylated. Nonetheless, SY209, because it produces few blue colonies on transformation with heteroduplex DNA, proved useful as a host for a biological assay of mismatch repair in mammalian cell extracts.

TABLE 2. Mismatch repair and DNA polymerase activity in HeLa cell extracts

Extract reaction conditions	Mismatch repair				DNA polymerase (% max [α-^{32}P]dCTP incorporation in DNA polymerase[c])
	Phenotype distribution of SY209 colonies[a]			% max[b]	
	No. blue	Total	% blue		
Complete	186	15,639	1.19	100	100
Without ATP, creatine kinase, and creatine phosphate	13	3,124	0.42	25	3
Without deoxynucleotides	11	3,158	0.35	18	63
With aphidicolin	16	4,805	0.33	16	2
With dideoxynucleotides	43	9,771	0.44	26	103
With dideoxynucleotides and without deoxynucleotides	9	2,988	0.3	13	NT[d]
With dideoxynucleotides and aphidicolin and without deoxynucleotides	8	3,950	0.2	3	NT
No extract	37	22,235	0.17	0	NT

[a] *E. coli* SY209 was transformed with heteroduplex DNA to yield ampicillin-resistant colonies of the given phenotype.
[b] Calculated by taking the percentage of blue colonies produced in each case, subtracting the percentage of blue colonies produced when no extract was used, and then normalizing to the value for the complete reaction.
[c] Calculated by subtracting the background counts and normalizing to the value for the complete reaction; 100% is equivalent to 24,034 cpm.
[d] NT, Not tested.

Activity in HeLa cell extracts. Heteroduplex DNA was added to cell extracts from HeLa cells that were prepared as described above. The incubation of the DNA in the extracts was carried out either under the conditions of the complete reaction as described above or with various experimental modifications. These modifications included the following: (i) no added ATP, creatine phosphate, or creatine kinase; (ii) no added deoxynucleotides; (iii) the addition of aphidicolin at 120 μM; (iv) the addition of dideoxynucleotides as follows: dideoxyguanosine triphosphate at 40 μM, dideoxyadenosine triphosphate at 40 μM, dideoxythymidine triphosphate at 80 μM, and dideoxycytidine triphosphate at 20 μM; (v) the addition of dideoxynucleotides at the concentrations given above in the absence of added deoxynucleotides; (vi) the addition of aphidicolin and dideoxynucleotides in the concentrations given above but without added deoxynucleotides; (vii) no incubation of the heteroduplex in the extract. The DNA was recovered from the extracts and used to transform SY209 to ampicillin resistance in the presence of X-gal, and the percentage of blue colonies was determined.

The results of these experiments are presented in Table 2. Preincubation of the heteroduplex in the complete extract prior to transformation of SY209 resulted in 1.19% blue colonies, whereas direct transformation of SY209 with the untreated heteroduplex yielded only 0.17% blue colonies. This sevenfold increase above the background value was statistically significant ($P < 0.001$), and it indicates that these extracts metabolize the heteroduplex DNA in some way. In the absence of exogenously added energy sources, the activity of the extracts above that of the background was reduced by about 75%. A similar effect was seen in the absence of added nucleotides, resulting in a reduction of extract activity by a factor of five. In the presence of aphidicolin or dideoxynucleotides, the activity of the extracts was again significantly reduced, by about 85 and 75%, respectively. When dideoxynucleotides were added in the absence of added deoxynucleotides, there was even less detectable activity, and when aphidicolin was also added with dideoxynucleotides in the absence of added deoxynucleotides, there was little if any effect above that of the background.

These results indicate that the process being measured in the extracts is energy dependent, is enhanced by the addition of deoxynucleotides, and is sensitive to aphidicolin and dideoxynucleotides, which are agents known to inhibit DNA polymerases (3, 16, 21, 22). Taken together, these results suggest that at least one aspect of the activity in the extracts involves DNA polymerization. To investigate this correlation further, extract reactions were set up under the same set of conditions as described above except that no unlabeled dCTP was added to any of the reactions but, instead, all reactions (including the reaction without deoxynucleotides) received [α-^{32}P]dCTP at a concentration of 20 μM (specific activity, 3,000 Ci/mmol). The proportion of radioactivity incorporated into trichloroacetic acid-insoluble material was determined in each case as one measure of DNA polymerization. This incorporation was compared with that found in the complete reaction, and in Table 2 the relative incorporation under each of the given conditions tested is expressed as a percentage of that in the complete reaction. There was little or no correlation between the effect of a given reaction modification on the mismatch repair assay as compared with the assay for dCTP incorporation, especially in the case of the addition of dideoxynucleotides alone (Table 2).

DNA recovered from the extract reactions carried out in the presence of [α-^{32}P]dCTP was analyzed by agarose gel electrophoresis and autoradiography. The results (data not shown) were consistent with the measurements of trichloroacetic acid-insoluble counts and demonstrated that the only labeled high-molecular-weight DNA species were those corresponding to the input DNA. No complex, reduced mobility forms that might indicate ongoing DNA replication were present, and no discrete degradation products were noted.

Recombination in HeLa cell extracts. An investigation of the activity of the HeLa cell extracts with regard to recombination was done as a control for the mismatch repair assay. These experiments are similar in design to those of Kucherlapati et al. (9) and are similar in concept to those of Darby and Blattner (2). Instead of forming heteroduplex molecules from p3AC-4 and p3AC-8, the two plasmids were both added directly to the extracts in either the circular or linear form. The DNA recovered after incubation in the extracts was used to transform SY209, as described above, and the percentage of blue colonies that was produced, which is indicative of the formation of a functional *supF* gene, was determined in each case (Table 3). No blue colonies were detected when circular plasmids were used to transform SY209, whether or not the mixed plasmids were preincubated in the extracts. When both plasmids were first linearized, very few colonies were produced at all, and none were blue. Only when p3AC-4, which was present in its circular form, was mixed with the linear form of p3AC-8 were any blue colonies detected, with there being 0.044% blue colonies after extract incubation and about eightfold fewer, or 0.0062% blue colonies, with no extract treatment. These results, like those reported previously (2, 9), suggest that there is some activity in the mammalian cell extracts which complements the RecA deficiency in SY209. Note that a much lower percentage of blue colonies was produced when the mixture of plasmids was used as a substrate as opposed to when heteroduplex prepared from the two mutant plasmids was used, and so the results with the heteroduplex DNA in the extracts cannot be attributed to recombination events alone.

TABLE 3. Recombination in HeLa cell extracts

DNA substrate[a]	Extract[b]	Phenotype distribution of SY209 colonies[c]		
		No. blue	Total	% blue
4 and 8	+	0	39,800	0
4 and 8	−	0	50,000	0
4 and 8, *Bam*HI	+	16	36,200	0.044
4 and 8, *Bam*HI	−	4	64,000	0.006
4, *Sca*I and 8, *Bam*HI	+	0	0	0
4, *Sca*I and 8, *Bam*HI	−	0	50	0

[a] Abbreviations: 4, p3AC-4; 8, p3AC-8.
[b] Either the complete extract and reaction conditions were used or the samples were incubated in buffer without HeLa cell extract.
[c] *E. coli* SY209 was transformed with DNA recovered from the extract reactions to yield ampicillin-resistant colonies of the given phenotype.

DISCUSSION

We set up a biological assay in which repair of heteroduplex plasmid DNA by human cell extracts could be subsequently detected by screening bacteria that were transformed with the plasmid DNA recovered from the extracts. The success of this assay depended on the construction of an *E. coli* strain in which mismatch repair and postreplicative plasmid recombination were minimal so that the background in the assay was low enough to measure the effect of the human cell extracts on the heteroduplex plasmid DNA. The HeLa cell extracts used in these experiments were essentially similar to the in vitro transcription extracts first described by Manley et al. (13), and they have been used in our laboratory for the study of transcription as well as mismatch repair (18).

In the assay the percentage of blue colonies produced by the transformation of SY209 with the heteroduplex DNA recovered from the extracts is taken as a measure of mismatch repair. The production of blue colonies depends on the conversion of one or both of the single base pair mismatches in the *supF* gene to the normal *supF* base pair. A single correction gives a blue colony if it is followed by plasmid replication, because in this case semiconservative replication generates a plasmid with a wild-type *supF* gene in addition to one with a mutant gene. Some events, however, may convert the mismatches to the mutant base pairs, but these events, because they would yield white colonies, are not specifically counted in this assay. These events appear among the many white colonies that arise from the introduc-

tion of unrepaired heteroduplex into the SY209 cells. Hence, the assay underestimates the frequency of match events. It might be expected that conversion of the mispair to either the wild-type or the mutant sequence is equally likely, as there were no particular differences between the strands of the heteroduplexes in these experiments, and so the actual frequency of repair events might be at least twice the observed value.

The results demonstrate that the HeLa cell extracts have a significant effect on the heteroduplex DNA. When the heteroduplex DNA was used to transform SY209 directly, only 0.17% of the colonies were blue. In contrast, incubation of the heteroduplex plasmid DNA in the extracts prior to transformation of SY209 produced 1.19% blue colonies, a sevenfold increase. Part of the assay, however, involved propagation of the DNA in *E. coli* before the final results could be determined, and so it is not known if all the steps involved in the repair of the mismatches are carried out and completed in the extracts. A significant part of the metabolism of the mismatched bases must occur in the extracts, because the results are dependent on the addition of ATP, creatine phosphate, creatine kinase, and deoxynucleotides to the extracts. The marked reduction in the activity of the extracts when these species were omitted indicates that some sort of energy-requiring enzymological process which needs a sufficient supply of DNA precursors is involved.

The exact nature of the activity in the extracts is not clear. Theoretically, some type of random nick-translation activity might be invoked to account for the repair of mismatches. Evaluation of the data, however, suggests that this explanation cannot account for all of the observed repair activity. The results show that the mismatch repair activity in the extracts is significantly reduced by the addition both of aphidicolin and dideoxynucleotides. Only aphidicolin, however, had an effect on the incorporation of labeled dCTP into the plasmid DNA. In fact, aphidicolin completely blocked [α-^{32}P]dCTP incorporation, whereas dideoxynucleotides had no measurable effect at all. The incorporation of [α-^{32}P]dCTP in these extracts is one measure of DNA polymerase activity. It is also an index of random nick-translation activity, because nick translation is a process that is associated with the incorporation of labeled nucleotides into DNA. Semiconservative DNA replication would also contribute to the results, but analysis of the reactions by gel electrophoresis showed that the pattern of complex, slowly migrating forms indicative of DNA replication was absent. Studies of in vitro replication have shown also that replication of plasmids, such as p3AC, which contain SV40 origins of replication, is dependent on exogenously added T-antigen protein, which was not added in these experiments. It is reasonable to assume, therefore, that the assay used to measure DNA polymerization in this study is essentially an assay for nick translation. Repair synthesis specifically associated with mismatch repair may also account for some [α-^{32}P]dCTP incorporation, but to the extent that it does contribute, the trivial explanation of random nick translation cannot be invoked. The data indicate that the presence of dideoxynucleotides in the extract reactions has no effect on the measured nick-translation activity but still reduces the apparent mismatch repair activity by about 75%. This suggests that some part of the observed mismatch repair activity that is sensitive to inhibition by dideoxynucleotides is unrelated to random nick translation. Similarly, in the absence of exogenously added deoxynucleotides, nick translation in the extracts was reduced to 63% of normal, while the apparent mismatch repair activity was just 18% of normal, again demonstrating an incomplete correlation between these activities in the extracts. Thus, although random nick translation cannot be ruled out as a partial explanation for the observed repair of mismatches, it does not appear to be associated with all of the mismatch repair activity in the extracts.

The activity that we observed may involve some nonspecific repair of mismatches which does not entail any particular strand selectivity. Bona fide mismatch repair, as it is understood in *E. coli*, involves a mechanism for strand discrimination so that repair can be directed toward the parental sequence. Based on the data from this study, we are unable to say whether or not the observed activity might involve some strand selectivity. The factors that affect strand selectivity in mammalian cells are not known for sure, but there has been one report (7) that when the substrate DNA is hemimethylated, there is an apparent bias toward the methylated strand in mismatch correction in vivo in monkey cells. We are currently investigating the effect of methylation on strand selection in vitro.

ACKNOWLEDGMENTS

We thank K. B. Low, E. Golub, R. D'Aquila, S. J. Baserga, and W. P. Summers for help with this study.

This study was supported by Public Health Service grants CA06519, CA13515, and CA16038 from the National Institutes of Health. P.M.G. was supported by Medical Scientist Training Program grant GM07205 from the National Institutes of Health.

ADDENDUM IN PROOF

Muster-Nassal and Kolodner (Proc. Natl. Acad. Sci. USA 83:7618–7622, 1986) have detected mismatch correction in cell-free extracts of *Saccharomyces cerevisiae*.

LITERATURE CITED

1. **Brown, E. L., R. Belagaje, M. J. Ryan, and H. G. Khorana.** 1979. Chemical synthesis and cloning of a tyrosine tRNA gene. Methods Enzymol. **68**:109–151.
2. **Darby, V., and F. Blattner.** 1984. Homologous recombination catalyzed by mammalian cell extracts in vitro. Science **226**:1213–1215.
3. **Dressler, S. L., and M. W. Lieberman.** 1983. Identification of DNA polymerases involved in DNA excision repair in diploid human fibroblasts. J. Biol. Chem. **258**:9990–9994.
4. **Folger, K. R., K. Thomas, and M. R. Capecchi.** 1985. Efficient correction of mismatched bases in plasmid heteroduplexes injected into cultured mammalian cell nuclei. Mol. Cell. Biol. **5**:70–74.
5. **Glickman, B. W., and M. Radman.** 1980. *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction. Proc. Natl. Acad. Sci. USA **77**:1063–1067.
6. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
7. **Hare, J. T., and J. H. Taylor.** 1985. One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repair. Proc. Natl. Acad. Sci. USA **82**:7350–7354.
8. **Herman, G. E., and P. Modrich.** 1981. *Escherichia coli* K-12 clones that overproduce *dam* methylase are hypermutable. J. Bacteriol. **145**:644–646.
9. **Kucherlapati, R. S., J. Spencer, and P. Moore.** 1985. Homologous recombination catalyzed by human cell extracts. Mol. Cell. Biol. **5**:714–720.
10. **Li, J. J., and T. J. Kelly.** 1984. Simian virus 40 DNA replication in vitro. Proc. Natl. Acad. Sci. USA **81**:6973–6977.
11. **Lindahl, T.** 1982. DNA repair enzymes. Annu. Rev. Biochem. **51**:61–87.
12. **Lu, A. L., S. Clark, and P. Modrich.** 1983. Methyl-directed repair of DNA base-pair mismatches *in vitro*. Proc. Natl. Acad.

Sci. USA 80:4639–4643.
13. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.** 1980. DNA-dependent transcription of adenovirus genes in a soluble whole cell extract. Proc. Natl. Acad. Sci. USA 77:3855–3859.
14. **Marinus, M. G., A. Poteete, and J. A. Arraj.** 1984. Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. Gene 28:123–125.
15. **Meyer, R. R., J. Glassberg, and A. Kornberg.** 1979. An *Escherichia coli* mutant defective in single-stranded binding protein is defective in DNA replication. Proc. Natl. Acad. Sci. USA 76:1702–1705.
16. **Miller, M. R., and D. N. Chinault.** 1982. The roles of DNA polymerases α, β, and γ in DNA repair synthesis induced in hamster and human cells by different DNA damaging agents. J. Biol. Chem. 257:10204–10209.
17. **Pukkila, P. J., J. Peterson, G. Herman, P. Modrich, and M. Meselson.** 1983. Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. Genetics 104:571–582.
18. **Read, G. S., and W. C. Summers.** 1982. *In vitro* transcription of the thymidine kinase gene of herpes simplex virus. Proc. Natl. Acad. Sci. USA 79:5215–5219.
19. **Sanger, E., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
20. **Sarkar, S., U. B. Dasgupta, and W. C. Summers.** 1984. Error prone mutagenesis detected in mammalian cells by a shuttle vector containing the *supF* gene of *Escherichia coli*. Mol. Cell. Biol. 4:2227–2230.
21. **Smith, C. A., and D. S. Okumoto.** 1984. Nature of DNA repair synthesis resistant to inhibitors of polymerase α in human cells. Biochemistry 23:1383–1391.
22. **Waqar, M. A., M. J. Evans, K. F. Manly, R. G. Hughes, and J. A. Huberman.** 1984. Effects of 2′,3′-dideoxynucleosides on mammalian cells and viruses. J. Cell Physiol. 121:402–408.
23. **Wobbe, C. R., F. Dean, L. Weissbach, and J. Hurwitz.** 1985. *In vitro* replication of duplex circular DNA containing the simian virus 40 DNA origin site. Proc. Natl. Acad. Sci. USA 82:5710–5714.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1990 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 265, No. 8, Issue of March 15, pp. 4718-4721, 1990
Printed in U.S.A.

DNA Damage Promotes Jumping between Templates during Enzymatic Amplification*

(Received for publication, September 21, 1989)

Svante Pääbo, David M. Irwin, and Allan C. Wilson

From the Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720

Pairs of templates and primers were designed so that only recombination events would lead to amplification via the polymerase chain reaction. This approach reveals that lesions such as breaks, apurinic sites, and UV damage in a DNA template can cause the extending primer to jump to another template during the polymerase chain reaction. By comparing sequences of amplification products that were determined directly or via bacterial cloning, it was shown that when the thermostable *Thermus aquaticus* DNA polymerase encounters the end of a template molecule, it sometimes inserts an adenosine residue; the prematurely terminated product then jumps to another template and polymerization continues, creating an *in vitro* recombination product. Consequently, amplification products from damaged templates such as archaeological DNA are made up of a high proportion of chimeric molecules. The illegitimate adenosine and thymidine residues in these molecules are detected when cloned molecules are sequenced, but are generally averaged out when the amplification product is sequenced directly. However, if site-specific lesions exist in template DNA or if the amplification is initiated from very few copies, direct sequencing also may yield incorrect sequences. The phenomenon of the "jumping polymerase chain reaction" can be exploited to assess the frequency and location of lesions in nucleic acids.

The polymerase chain reaction (1) is becoming widely used in molecular biology because it can detect and amplify a few or even single copies of a DNA segment. Polymerase chain reaction is of particular value for the study of DNA in single cells (2), forensic samples (3), archaeological remains (4), and museum specimens (5). However, in the latter cases, the vast majority of DNA molecules present have been shown to be damaged (6). It may be speculated that these damaged molecules can contribute in various ways to the population of molecules that make up the final amplification product. Similarly, when amplifications are initiated from single template molecules, damage present in the template molecule could influence the results.

Such considerations prompted us to investigate the effects of various types of damage in the template DNA on the polymerase chain reaction as performed with the thermostable *Thermus aquaticus* (*Taq*) DNA polymerase, particularly with respect to insertions of incorrect bases and the generation of recombination products *in vitro*. This investigation also led us to design a polymerase chain reaction method that may prove valuable for detecting and measuring DNA damage.

* This work was funded by a European Molecular Biology Organization Fellowship, a Canadian Medical Research Council Fellowship, and grants from the National Science Foundation and the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

EXPERIMENTAL PROCEDURES

Enzymatic amplifications were performed in 25-μl reaction mixtures containing 67 mM Tris/Cl, pH 8.8, 2 mM $MgCl_2$, 250 μM concentration each of dATP, dCTP, TTP, and dGTP and 1.25 units of *T. aquaticus* DNA polymerase (Perkin-Elmer-Cetus). In the experiments where ancient DNA was used, 2 μg/ml bovine serum albumin (Sigma, fraction V) was added to the reactions. Agarose gel electrophoresis and direct sequencing were performed as described by Pääbo *et al.* (4). Unless otherwise stated, 2 ng of each template construct was present in 25-μl reactions. Primers used were: the M13 "universal" primer (7), D3E and D18X (4), L14841 (8), H14876 (5'-CGCTGCAGAATAGGCCTGTTAGGATTTG-3'), L16175 (5'-GCAAGCTTAGTACATAAAAACCCAATCCA-3'), DI5 (5'-AAGATCTTTGAGAGATGTGA-3'), and SP1 (5'-TACCCGGGGCTGAGCCATCACTCAA-3'). Forty cycles of polymerase chain reaction were performed as follows: denaturation at 92 °C for 40 s, annealing at 52 °C for 1 min, and extension at 72 °C for 1 min. Sonication was done in a model H-IL sonicator (Ultrasonics) for 5 × 30 s on ice, which produced fragments ranging in size from 100–1500 nucleotides. Depurination was performed by incubating the DNA in 0.1 N HCl for 5 min at room temperature. Tris Cl, pH 8.8, was then added to a final concentration of 0.1 M. UV irradiation was performed by illumination of the DNA sample on a UV table (UVP) for 1 min. The depurinated and UV-irradiated templates showed no evidence of degradation when analyzed by agarose gel electrophoresis.

RESULTS

Jumping Induced by Breaks in Template—In order to assess the tendency for forming recombinant amplification products, polymerase chain reaction was carried out from a set of two template molecules (Fig. 1). Template I was a cDNA containing the complete protein-coding sequence of cow lysozyme type 2b (clone λcBL42, Ref. 9). Template II was a genomic clone of a cow lysozyme type 3 gene (clone pLI, Ref. 10) which contains exons 2, 3, and 4 as well as introns 2 and 3. Within exon 2, template II differs from template I only by having cytidine residues at positions 290 and 293 (Fig. 2). These two template molecules were mixed with one primer specific to exon 1 (*primer A* in Fig. 1), which occurs in template I but not in template II, and another primer (*primer B* in Fig. 1) which is specific for intron 2 and occurs in template II but not in template I. Thus, no template molecule containing both of the primer sites was added to the reactions, and exon 2 is the only region located 3' to the primers where sequence similarity between the two template molecules exists. After 40 cycles of polymerase chain reaction, no specific product could be detected when the amplification reaction was analyzed by ethidium staining of an agarose gel (Fig. 1, *bottom left, lane 4*).

When template I was cut with *Pst*I and/or template II with *Sau*3AI, an amplification product of the expected size for an *in vitro* recombination product between the two templates

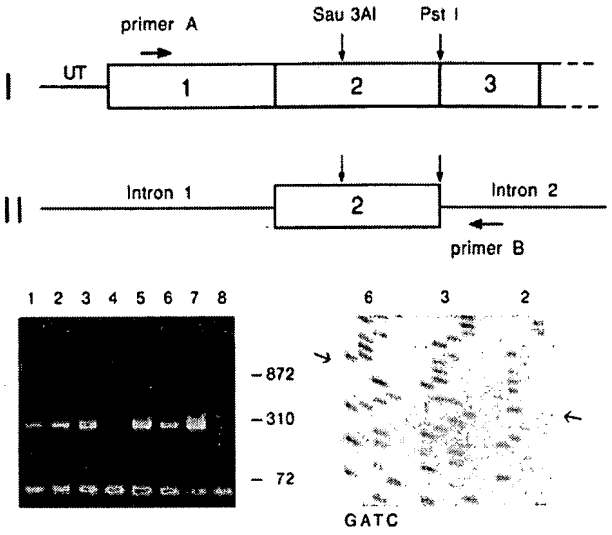


FIG. 1. **Jumping polymerase chain reaction illustrated with two lysozyme clones.** *Above*, parts of the two lysozyme clones as well as primers and restriction sites used are schematically illustrated. *Primer A*, specific for exon 1, is DI5, and *primer B*, specific for intron 2, is SP1 (see "Experimental Procedures" for details). *Numbers* indicate protein-coding exons, and *UT* denotes the 5'-untranslated region of the cDNA. *Below at the left*, agarose gel electrophoresis of amplification reactions are shown. The templates used were: *1*, template I *Pst*I-cut, template II unrestricted; *2*, template I unrestricted, template II *Sau*3AI-cut; *3*, template I *Pst*I-cut, template II *Sau*3AI-cut; *4*, both templates unrestricted; *5*, templates sonicated; *6*, templates UV-irradiated; *7*, templates depurinated; *8*, no template. The lower molecular weight bands are dimers of primers. The migration positions of molecular size markers are indicated in numbers of base pairs. *Below at the right*, part of the direct sequencing reaction around the *Sau*3AI site is shown for the UV-irradiated template (*lanes* marked *6*), templates I and II cut with *Pst*I and *Sau*3AI, respectively (*lanes* marked *3*) and uncut and *Sau*3AI-cut, respectively (*lanes* marked 2). The *arrows* point to position 249.

was generated (Fig. 1). Upon direct sequencing (11), this product was shown to have the expected composite sequence where axons 1 and 2 were joined. In addition, at positions 290 and 293 where the two template sequences differ, direct sequencing of the amplification product yielded an equal mixture of the two sequences when both templates had been cut. When only one of the templates were cut, the sequence of that template was present at these sites. These results show that *in vitro* recombination events occur readily when breaks exist in the template molecules and that this phenomenon is dependent on sequence similarity between the two extended primers, since restriction of templates I and II with *Sau*3AI and *Pst*I, respectively, did not yield any relevant amplification product.

Insertion of Adenosine Residues—Direct sequencing of the amplification products showed that when template II had been cut with *Sau*3AI and template I with *Pst*I, an adenosine residue was present in about equal frequency with the expected thymidine residue at position 249 (Fig. 1, *lower right*, *lanes 3*). This represents the position at which template II had been cut with *Sau*3AI. When only template II was cut, only an adenosine residue was detected. This demonstrates that *in vitro* recombination between template molecules during polymerase chain reaction can cause unambiguous and incorrect sequences to be obtained. In experiments where the *Pst*I site was cut either alone or in conjunction with the *Sau*3AI, no aberrant bases were detected by direct sequencing.

The polymerase chain reaction product generated from the *Pst*I- and *Sau*3AI-cut templates was cut with *Pst*I and *Bgl*II and cloned in the vector M13mp19 in order to determine the sequence at the *Sau*3AI site. Nine clones were sequenced for the 234 bp[1] from the DI5 primer site to the *Pst*I restriction site (Fig. 2). Eight of these clones had cytosine residues at positions 290 and 293, which indicates that they originate from template II. Seven out of these eight clones carry a thymidine residue at position 249 instead of the expected adenosine. This confirms the results of the direct sequencing and shows that *Taq* polymerase can insert adenosines when it reaches the end of the template DNA strand. In addition to these substitutions, the polymerase chain reaction products display four substitutions where 2 cytosine, 1 adenosine, and 1 thymidine residues are gained in the 2106 bp sequenced. This reflects the normal error frequency that we and others (12, 13) observe when polymerase chain reaction products are cloned.

DNA Damage Induces Jumping—In order to determine whether random damage to the template DNA could induce jumping polymerase chain reaction, aliquots of the two unrestricted template DNA preparations were subjected to sonication, depurination by acid, and UV irradiation, respectively. In all cases, the resulting templates were able to generate a recombinant polymerase chain reaction product (Fig. 1, *bottom left*) that upon direct sequencing proved to contain the expected sequence.[2]

The products were cloned, and a total of 27 clones were sequenced (Fig. 2). The majority of these clones carry C at positions 290 and 293. This is expected, since the probability that a particular sequence would be originating from template II should increase with its proximity to primer B if the position within exon 2 at which the jumping between templates occurs is random. In one case (clone jU26), the jumping seems to have occurred between positions 290 and 293. Also, the number of substitutions seems to be elevated in intron 2. However, no net increase of adenosine or thymidine residues can be seen.

Jumping PCR Promoted by Ancient DNA—To elucidate whether the jumping phenomenon occurs in amplifications from ancient DNA, we amplified a region of the mitochondrial control region from the DNA extracted from the 4000-year-old mummified liver of an Egyptian priest. This DNA segment had previously been amplified and directly sequenced from this individual and shown to differ at two positions from a published human mitochondrial sequence (6). When the amplification product was cloned and four clones were sequenced (Fig. 3), one proved to be identical with the directly determined sequence. The other three clones displayed the same two differences from the reference sequence but in addition carried substitutions where cytidine residues had been replaced by thymidine residues in six cases and a guanosine residue had been replaced by an adenosine residue in one case. This distribution of changes is consistent with the nucleotide composition of this DNA segment as well as with the observation that in this as well as other archaeological DNA samples pyrimidines are predominantly modified or missing (6). At these sites, adenosines may have been inserted on the opposite strand by two different mechanisms: (*a*) during polymerization without jumping and (*b*) by eliciting jumping. In both cases, only the changes occurring opposite cytidine residues would lead to incorrect nucleotides being incorporated. Thus, all of the substitutions detected would be

[1] The abbreviation used is: bp, base pair(s).

[2] Also, high template DNA concentrations (in the order of 1 μM) may cause high frequencies of jumping in a system similar to the one used here (M. Dutreix, unpublished observation).

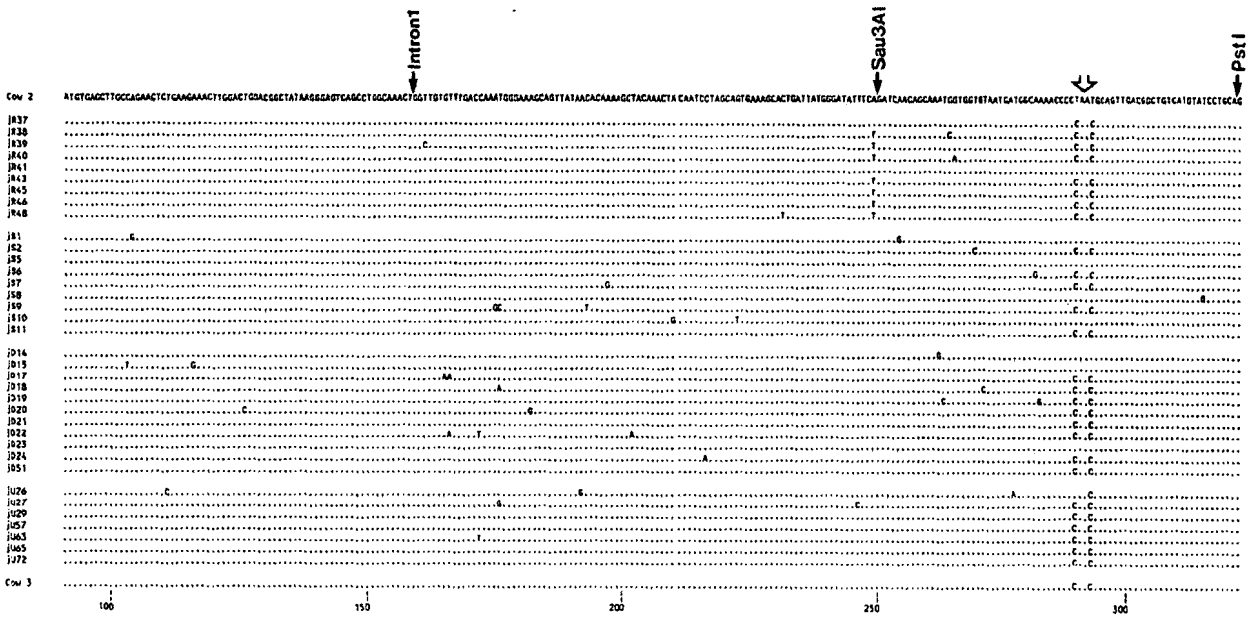


FIG. 2. DNA sequences of cloned amplification products resulting from the jumping polymerase chain reaction. The templates were the cow lysozyme 2 cDNA and lysozyme 3 genomic clones in Fig. 1 which had been restricted (*jR*), sonicated (*jS*), depurinated (*jD*), and UV-irradiated (*jU*), respectively. *Arrows* point to the positions at which indicated restriction enzymes cleave the DNA strand shown and where intron 1 exists in the genomic clone. The *open arrow* points to positions 290 and 293.

```
Reference  ACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAA
Direct Seq .............T.........................C.....
Clone 1    .............T.........................C.....
Clone 2    ....T.......TT....T...T.................C.....
Clone 3    ...A.........T.........................C.....
Clone 4    .............T.T..T.....................C.....
```

FIG. 3. DNA sequences of a 45-bp segment of the mitochondrial control region from a 4000-year-old Egyptian mummy. The published human sequence (above; Ref. 22) is compared to the directly sequenced amplification product (identical with results in Ref. 6) and four clones isolated from the product of the amplification reaction.

transitions at positions occupied by guanosine and cytidine residues on the strand whose sequence appears in Fig. 3.

The fact that some clones contain numerous adenosine/thymidine substitutions in conjunction with the fact that "jumping" occurring at modified thymidines is expected to predominate indicates that the frequency of modified bases and strand breaks in the ancient DNA is extremely high, causing several "jumping events" or misincorporations to occur per molecule amplified. Since pyrimidines are predominantly damaged in ancient DNA (6) and since adenosines inserted at sites opposite modified thymidine residues will not cause substitutions, a quantitative estimate of the frequency of damage can be made for cytosine residues only. For these residues, the sequences in Fig. 3 indicate that a minimum of 1 residue in 11 (7 cytosine residues out of 76 sequenced) causes the misincorporation of an adenosine and thus is damaged or absent from the template DNA.

An Assay for DNA Damage—To demonstrate directly that jumping polymerase chain reaction occurs when amplification is performed from ancient DNA, a copy of the mitochondrial sequence cloned in M13 was used together with approximately 1 μg of the 4000-year-old DNA in a "jumping polymerase chain reaction assay" (Fig. 4). The primers for amplification were the M13 universal primer (*A*) and a primer (*B*) located in the mitochondrial control region outside the region cloned. Thus, any generation of a product by these primers must be due to an intermolecular amplification initiated by recombination between the mitochondrial segment present in the clone and corresponding sequences prepared from the ancient Egyptian individual. As a control, two additional primers (*C* and *D*) specific for a segment of the mitochondrial cytochrome *b* gene were added to the reaction. They are expected to give rise to a conventional, intramolecular amplification product of 97 bp and serve as an internal control for the amounts of noncloned template DNA added as well as for the overall efficiency of the reaction. Similar reactions were set up with total DNA prepared from contemporary autopsy material. The intramolecular amplification proved to be dramatically more efficient than the intermolecular amplification. In order to achieve comparable intensities of the two products, it was necessary to perform 15 cycles of polymerase chain reaction with primers intended only for the jumping polymerase chain reaction and then add the primers for the intramolecular amplification prior to performing 25 additional cycles. As can be seen in Fig. 4, the relative amount of the intermolecular polymerase chain reaction is appreciably greater in amplifications performed from the ancient DNA extract than from the contemporary DNA preparation. This demonstrates the existence of many more lesions inducing the partially extended amplification products to jump from one template molecule to another in the ancient DNA than in the modern DNA.

DISCUSSION

Others have shown that DNA fragments can be fused for construction purposes (14) by the approach that we designate as the jumping polymerase chain reaction, that "shuffle clones" are sometimes detected when amplification products are cloned (15), and that *Taq* polymerase shares with other prokaryotic and eukaryotic DNA polymerases the propensity for inserting adenosines (16) when no template base is present. Furthermore, the lack of any 3′-5′ exonuclease activity in *Taq* polymerase (17) ensures that the additional base is not

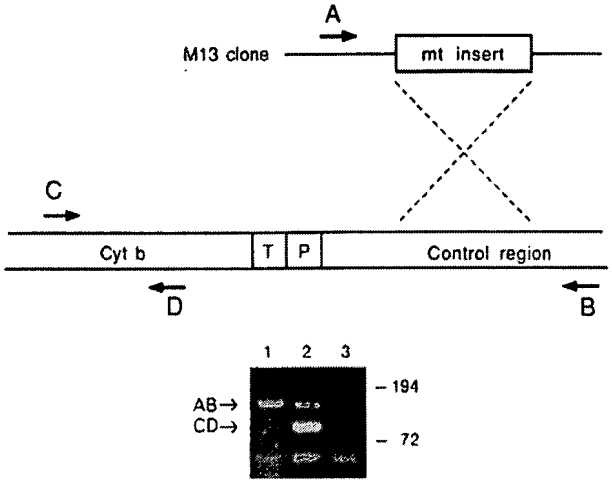


FIG. 4. **Demonstration of damage in ancient samples of DNA by the jumping polymerase chain reaction.** *Above*, a schematic illustration showing the cloned mitochondrial sequence; the universal M13 primer (*primer A*) and the mitochondrial *primer B* (L16175) used for the intermolecular amplification; *primers C* (L14841) and *D* (H14876) used as an internal control; the mitochondrial DNA analyzed for lesions by jumping polymerase chain reaction where the control region and the genes for $tRNA^{Thr}$, $tRNA^{Pro}$, and cytochrome *b* are indicated. *Below*, agarose gel electrophoresis of amplification reactions from the 4000-year-old liver DNA (*1*), total liver DNA from contemporary autopsy material (*2*), and no uncloned template DNA (*3*). *AB* denotes the 160-bp intermolecular amplification products, and *CD* the 97-bp intramolecular product. The identities of the former products were confirmed by direct sequencing. The migration positions of molecular size markers are indicated in numbers of base pairs.

removed after its incorporation into the nascent strand. However, our observation that no addition of adenosines occurs at the *PstI* site indicates that this phenomenon is dependent not only on properties of *Taq* polymerase but also on the sequence context and underscores the need to sequence DNA constructs generated by polymerase chain reaction-induced joining of molecules.

From the experiments described above, it is clear that if amplifications are initiated from single or very few template molecules, any damage inducing the jumping polymerase chain reaction may be reflected in a large fraction or even all of the molecules making up the final amplification product. Furthermore, if site-specific breaks (*e.g.* induced by the HO endonuclease in the yeast *MAT* locus; Ref. 18) or nicks (*e.g.* induced by topoisomerase I; Ref. 19) occur in the DNA prepared from an organism, almost all molecules in the amplification product may be of incorrect sequences. However, if the damage in the template DNA is randomly distributed but so extensive that few or no intact molecules containing both primer sites exist, the jumping polymerase chain reaction will allow amplifications of longer DNA segments than are actually present in the sample. This is because the primers during the first cycles of amplification are extended on different templates until one or both of them reach the reciprocal priming site. After this initial lag phase, the length of which is proportional to the amount of damage present, a molecule containing both primer sites has been created and an exponential amplification will ensue (20). However, a substantial part of the product molecules will be rearranged and contain inserted adenosines. In fact, even the majority of the molecules may carry such abnormalities (Fig. 3), but when the product is directly sequenced, only a consensus sequence reflecting the unperturbed template sequence will be scored (21).

A serious concern is that if the amplification product is not sequenced but rather typed for particular alleles with allele-specific oligonucleotides, *in vitro* recombination phenomena may go undetected. In fact, new specificities may be created by the jumping polymerase chain reaction combining DNA segments from different alleles or loci. In particular, this may be the case when amplifications are performed from nuclear genes in heterozygous individuals, from genes belonging to multigene families such as the major histocompatibility complex, or from extracts that contain DNA from more than one individual or species.

The experimental design illustrated in Fig. 4 can be used to assess the amount of damage present in a particular DNA segment. The lesions can in many cases be determined to the exact base at which they occur by the cloning and sequencing of the product; it is to be expected that adenosine or thymidine residues should be present at sites of damage (*cf.* Fig. 3). However, these applications of the jumping polymerase chain reaction require further investigation into the molecular nature of the DNA modifications that cause *Taq* polymerase to stall or fall off its template. Also, finding conditions that promote stalling or falling off the template may improve the sensitivity of this assay system.

Acknowledgments—We are grateful to W. Kelley Thomas, Thomas D. Kocher, Ellen M. Prager, Thomas J. White, Charles Radding, Marie Dutreix, and David H. Gelfand for valuable discussions and comments on the manuscript.

REFERENCES

1. White, T. J., Arnheim, N., and Erlich, H. A. (1989) *Trends Genet.* **5**, 185–189
2. Li, H., Gyllensten, U. B., Cui, X., Saiki, R. K., Erlich, H. A., and Arnheim, N. (1988) *Nature* **335**, 414–417
3. Higuchi, R., von Beroldingen, C. H., Sensabaugh, G. F., and Erlich, H. A. (1988) *Nature* **332**, 543–546
4. Pääbo, S., Gifford, J. A., and Wilson, A. C. (1988) *Nucleic Acids Res.* **16**, 9775–9787
5. Thomas, R. H., Schaffner, W., Wilson, A. C., and Pääbo, S. (1989) *Nature* **340**, 465–467
6. Pääbo, S. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 1939–1943
7. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
8. Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., and Wilson, A. C. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6196–6200
9. Irwin, D. M., and Wilson, A. C. (1989) *J. Biol. Chem.* **264**, 11387–11393
10. White, R. T., Dobson, D. E., Wilson, A. C., Jollès, J., Schoentgen, F., and Jollès, P. (1983) *Fed. Proc.* **42**, 1971
11. Gyllensten, U. B., and Erlich, H. A. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 7652–7656
12. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239**, 487–491
13. Tindall, K. R., and Kunkel, T. A. (1988) *Biochemistry* **27**, 6008–6013
14. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res.* **16**, 7351–7367
15. Shuldiner, A. R., Nirula, A., and Roth, J. (1989) *Nucleic Acids Res.* **17**, 4409
16. Clark, J. M. (1988) *Nucleic Acids Res.* **16**, 9677–9686
17. Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. (1989) *J. Biol. Chem.* **264**, 6427–6437
18. Nasmyth, K. (1983) *Nature* **302**, 670–676
19. Stewart, A. F., and Schütz, G. (1987) *Cell* **50**, 1109–1117
20. Pääbo, S., Higuchi, R. G., and Wilson, A. C. (1989) *J. Biol. Chem.* **264**, 9709–9712
21. Pääbo, S., and Wilson, A. C. (1988) *Nature* **334**, 387–388
22. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., and Young, I. G. (1981) *Nature* **290**, 457–465

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1988 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 263, No. 14, Issue of May 15, pp. 6829-6835, 1988
Printed in U.S.A.

Mispair Specificity of Methyl-directed DNA Mismatch Correction *in Vitro**

(Received for publication, October 9, 1987)

Shin-San Su, Robert S. Lahue‡, Karin G. Au§, and Paul Modrich

From the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710

To evaluate the substrate specificity of methyl-directed mismatch repair in *Escherichia coli* extracts, we have constructed a set of DNA heteroduplexes, each of which contains one of the eight possible single base pair mismatches and a single hemimethylated d(GATC) site. Although all eight mismatches were located at the same position within heteroduplex molecules and were embedded within the same sequence environment, they were not corrected with equal efficiencies *in vitro*. G-T was corrected most efficiently, with A-C, C-T, A-A, T-T, and G-G being repaired at rates 40–80% of that of the G-T mispair. Correction of each of these six mispairs occurred in a methyl-directed manner in a reaction requiring *mutH*, *mutL*, and *mutS* gene products.

C-C and A-G mismatches showed different behavior. C-C was an extremely poor substrate for correction while repair of A-G was anomalous. Although A-G was corrected to A-T by the *mutHLS*-dependent, methyl-directed pathway, repair of A-G to C-G occurred largely by a pathway that is independent of the methylation state of the heteroduplex and which does not require *mutH*, *mutL*, or *mutS* gene products. Similar results were obtained with a second A-G mismatch in a different sequence environment suggesting that a novel pathway may exist for processing A-G mispairs to C-G base pairs.

As judged by DNase I footprint analysis, MutS protein is capable of recognizing each of the eight possible base-base mismatches. Use of this method to estimate the apparent affinity of MutS protein for each of the mispairs revealed a rough correlation between MutS affinity and efficiency of correction by the methyl-directed pathway. However, the A-C mismatch was an exception in this respect indicating that interactions other than mismatch recognition may contribute to the efficiency of repair.

The fidelity of DNA replication in *Escherichia coli* is enhanced 100–1000-fold by a postreplication mismatch correction system (reviewed in Refs. 1–4). The requisite strand specificity is provided in this system by the methylation state of d(GATC) sites, with repair directed to the unmethylated strand in hemimethylated DNA (5–8). Methyl-directed mismatch correction in this organism requires the products of *mutH*, *mutL*, and *mutS* genes, as well as DNA helicase II and single-stranded DNA binding protein (SSB) (6, 9–11). The protein products of the *mutH*, *mutL*, and *mutS* genes have recently been purified to near homogeneity in biologically active form (12–14). The MutS protein has been shown to recognize several mismatched base pairs (12), while analysis of the MutH protein suggests that it functions in strand discrimination by incising the unmethylated DNA strand at d(GATC) sites (14). No activity has been assigned to the isolated MutL protein.

Transfection of *E. coli* with artificially constructed heteroduplexes has demonstrated that the different mismatches are subject to correction with different efficiencies *in vivo* (15–17). G-T, A-C, G-G, and A-A mismatches are typically subject to efficient repair. A-G, C-T, T-T, and C-C are weaker substrates, but well repaired exceptions exist within this class. The elements of mismatch structure and features of the repair system that determine repair efficiency are not understood. However, it is thought that the sequence environment of the mismatch may be an important factor (4, 17), and the affinity of MutS protein for several different mispairs has been shown to vary (12). In this paper, we have utilized an *in vitro* mismatch repair assay (7) to examine repair efficiencies of the eight possible single base pair mismatches which were constructed in such a way that each is embedded within the same DNA sequence environment. We also show that MutS protein is capable of recognizing all eight mismatches and that there is some correlation between the affinity of MutS protein for the mispairs and their correction efficiencies.

EXPERIMENTAL PROCEDURES

E. coli Strains—AB1157 (*thr-1 leu-6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1 kdgK51 proA2 his-4 argE3 str-31 tsx-33 supE44*) was from V. Burdett (Duke University), RK1517 (as AB1157 but *mutS*::Tn*5*) was from R. Kolodner (Dana-Farber Cancer Institute). Strain RS5033 (HfrH *metB1 rel-1 str-100 azi-7 lacMS286 thi dam-4 φ80dIIlacBK1*) was provided by E. B. Konrad. JJ119 (HfrH *metJam185*) was from R. Greene (Duke University), and JJ119 harboring plasmid pTP166, which carries the *dam* gene under *tac* control (18), has been described previously (8).

DNA Preparations and Mismatch Repair Assay—Bacteriophage f1MR1, which contains a single d(GATC) site at position 216, was derived from f1h, (8) as summarized in the legend to Table I. A set of derivatives of f1MR1 allowing preparation of the eight possible base-base mismatches was constructed by sequential, site-directed oligonucleotide mutagenesis according to Kramer *et al.* (15). The sequences of oligonucleotides used for mutagenesis and the corresponding f1MR mutants are shown in Table I. Mutant f1MR phages were identified by plaque hybridization (19) using appropriate oligonucleotides as probes, and mutant sequences were confirmed by dideoxy sequencing (20).

Phage f1G18 (12) is a derivative of f1R229 (21) containing a T to G transversion mutation at position 5620, converting the *Eco*RI site (G-A-A-T-T-C) of the virus to a *Bsm*I site (G-A-A-T-G-C). Deriva-

* This work was supported by Research Grant GM23719 from the National Institute of General Medical Sciences. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by a Damon Runyon-Walter Winchell Cancer Fund Postdoctoral Fellowship.

§ Supported by a fellowship from The Ladies Auxiliary to the Veterans of Foreign Wars. Currently a Syntex Fellow of the Life Sciences Research Foundation.

tives of f1G18 containing only one d(GATC) site (position 216, $f1G18_1$) or no d(GATC) sites ($f1G18_0$) were constructed by exchange of the 0.9-kilobase *Hgi*AI-*Ban*II fragment (coordinates 4743–5649) from phage f1G18 for the corresponding fragment of phage $f1h_1$ or $f1h_0$ (8).

Replicative form DNAs methylated at the d(GATC) site at position 216 were isolated from strain JJ119, methylated single-stranded DNAs were isolated from virions propagated in strain JJ119(pTP166) (8), and unmethylated DNAs were isolated from strain RS5033. DNA heteroduplexes were prepared and mismatch correction in cell-free extracts determined as described previously (7, 8), except that mismatch repair reactions were supplemented with 0.08 M KCl.

DNase I Footprinting—A 143-base pair *Hae*II-*Ban*I fragment (coordinates 5572–5714) was isolated from each f1MR heteroduplex after 3′-end labeling of the *Ban*I terminus using Klenow DNA polymerase I (United States Biochemicals) and [α-^{32}P]dATP (3000 Ci/mmol, Du Pont-New England Nuclear). Footprinting reactions (15 μl) contained 0.02 M Tris-HCl (pH 7.6), 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 1 mM dithiothreitol, 0.1 mM EDTA, 125 fmol of the end-labeled *Hae*II-*Ban*I fragment, and Fraction IVb MutS protein (12) as indicated (22). After 10 min at 30 °C, DNase I (1.7 ng in 1 μl) was added and incubation was continued for 10 s (in the absence of MutS protein) or 20 s (in the presence of MutS protein) before quenching (22). Samples were analyzed on 8% polyacrylamide sequencing gels.

RESULTS

Construction of Heteroduplexes in Which the Eight Single Base Pair Mismatches Reside in a Common Sequence Environment— We have previously described the construction of a 6.4-kilobase f1 heteroduplex containing a single d(GATC) sequence and a G-T mispair residing within overlapping recognition sites for restriction endonucleases *Xho*I and *Hin*dIII (8). As shown in Fig. 1, this approach allows independent assay of mismatch correction on either DNA strand. We have shown previously that the state of methylation of the single d(GATC) site is sufficient to direct the correction of the G-T mismatch which is located at a distance of 1024 base pairs as measured along the shorter path (8).

Starting from phages f1MR1 and f1MR3, which were used to construct the G-T heteroduplex mentioned above, we have prepared additional f1MR derivatives (Table I) that permit construction of a set of heteroduplexes representing the eight possible single base pair mismatches. Each of these heteroduplexes contains a single d(GATC) site at position 216, and in each case the mismatch is located at the same position (coordinate 5632). Moreover, as shown in Table II, each mispair in the set of eight heteroduplexes is located within overlapping restriction endonuclease recognition sites, and in every case the four base pairs on either side of the mismatch are identical. The generation of overlapping restriction sites was accomplished while maintaining the local sequence environment by variation of the fifth base pair on either side of the mismatch. This set of heteroduplexes allows correction on either DNA strand to be directly determined under conditions where effects of sequence environment on repair efficiency are minimized.

Efficiency of in Vitro Correction Is Dependent on the Mismatch—Each of the eight heteroduplexes was tested for efficiency of *in vitro* repair in each of the two hemimethylated configurations. Table II compares the efficiency of correction of the eight mismatches as determined by initial rate measurements, and Fig. 2 illustrates the behavior of the several repair classes in the restriction assay. As shown in Table II, six of the heteroduplexes (G-T, A-C, C-T, A-A, T-T, and G-G) were subject to methyl-directed repair in cell-free extracts of *E. coli*. As judged by initial rate determinations, the G-T mismatch was the preferred substrate, with A-C, C-T, A-A, T-T, and G-G corrected at rates of 40–80% of that observed for the G-T mispair. These differences in repair efficiencies do not reflect the presence of inhibitors in some DNA preparations since experiments in which each heteroduplex was competed with the G-T substrate resulted in the same hierarchy of correction (not shown).

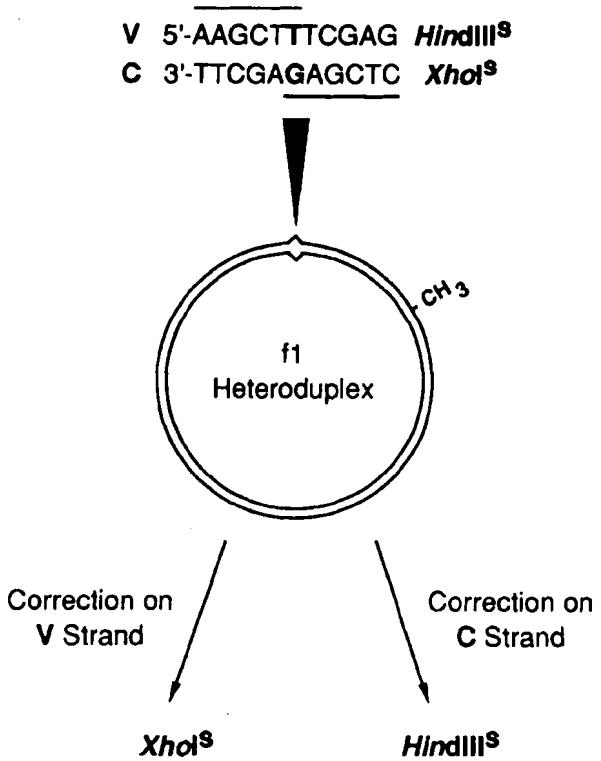


FIG. 1. **Heteroduplex assay for *in vitro* mismatch correction.** Each heteroduplex used in this study is a 6440-base pair, covalently closed circular molecule containing a single base pair mismatch at position 5632. The presence of the mispair within overlapping restriction endonuclease sites allows the strand specificity of correction to be determined, as illustrated here for a G-T mismatch within overlapping *Hin*dIII and *Xho*I sites. The heteroduplexes also contain a single d(GATC) sequence at position 216, 1024 base pairs from the mismatch. The states of methylation of this d(GATC) site were controlled as described under "Experimental Procedures." *C* and *V* designate complementary and viral strands of the f1 DNA molecule.

Repair of either hemimethylated configuration of each of these six heteroduplexes occurred preferentially on the unmethylated DNA strand. As shown in Fig. 2 and Table II, repair bias exceeded 6:1 (unmethylated:methylated) in all cases except for the C-T heteroduplex that was modified on the viral strand. In this case repair was biased to the unmethylated strand by a factor of only 3.5 to 1, a value which may be due in part to the fact that this preparation of heteroduplex was contaminated at the 20% level by molecules which were unmethylated on either DNA strand (not shown). This analysis also demonstrated that in the case of G-T, A-C, C-T, A-A, T-T, and G-G heteroduplexes, the hemimethylated configuration bearing the methyl group on the complementary strand was a superior substrate to that in which the modification resided on the viral strand. The possible significance of this observation will be considered below.

We have previously demonstrated that as observed *in vivo*, *in vitro* correction of a G-T and an A-C mismatch is dependent on the presence of *mutH*, *mutL*, and *mutS* gene products (7). As shown in Table III, correction of the six mispairs described above (G-T, A-C, C-T, A-A, T-T, and G-G) is highly dependent on the presence of the *mutS* gene product in crude

TABLE I

Construction of f1 derivatives for heteroduplex preparation

As described previously (8), bacteriophage f1MR1, a derivative of $f1h_1$ containing a single d(GATC) site at position 216, was constructed by elimination of the *Cla*I site at position 6040 of $f1h_1$ followed by insertion into the *Eco*RI site at position 5616 of the 27-base pair synthetic duplex

```
                    *
V 5'-AATTGCTAGCAAGCTTTCGAGTCTAGA
      CGATCGTTCGAAAGCTCAGATCTTTAA-5' C.
```

Phage f1MR3 containing a single T to C base change at position 5632 was derived from f1MR1 by oligonucleotide mutagenesis (8, 15). With f1MR1 and f1MR3 as precursors, a set of additional f1MR phages has been constructed for this study. The *underlined* nucleotide in the column of viral strand sequence indicates the base change for each step. The fourth column shows the oligonucleotide used to modify the parent phage, with the *underlined base* being noncomplementary to the parental viral strand. An * indicates position 5632, the location of mismatches within heteroduplexes constructed with these molecules (Table II). V, viral strand; C, complementary strand.

f1 mutant	Viral strand sequence (5′ → 3′)	Parent phage	Mutagenic oligonucleotide (5′ → 3′)	Marker
	*			
f1MR1	AAGCTTTCGAG			*Hin*dIII
f1MR3	AAGCTCTCGAG	f1MR1	ACTCGAGAGCTTGC	*Xho*I
f1MR4	AAGCTTTCGAT	f1MR1	TCTAGAATCGAAAG	*Hin*dIII
f1MR5	AAGCTATCGAT	f1MR4	GAATCGATAGCTTG	*Cla*I
f1MR6	AAGCTTTCGAC	f1MR1	TTCTAGAGTCGAAA	*Hin*dIII
f1MR7	AAGCTGTCGAC	f1MR6	GAGTCGACAGCTTG	*Sal*I
f1MR8	CAGCTCTCGAG	f1MR3	AGAGCTGGCTAGCA	*Xho*I
f1MR9	CAGCTGTCGAG	f1MR8	GACTCGACAGCTGG	*Pvu*II

TABLE II

Rate of in vitro repair depends on the mismatch

Covalently closed, circular heteroduplexes containing the eight possible base pair mismatches were prepared (7) in both hemimethylated configurations using the phage DNAs shown. All eight mismatches (**bold type**) were located at position 5632 with the four base pairs on either side of this position being the same in each case. Variation at the fifth base pair (*lower case*) on either side of the mismatch allowed placement of mispairs within overlapping restriction sites that serve as strand markers. With the exception of the nature of the mismatch and this fifth position variation, all heteroduplexes were identical. Complementary and viral strands are designated C and V, with the methylation state of a strand indicated by + or −. Bias indicates the relative efficiency of repair on the two DNA strands (unmethylated/methylated) as determined after 60 min reaction. Mismatch correction was determined in extracts of AB1157 (10 mg/ml protein) as described under "Experimental Procedures," except that reactions were sampled at 5-min intervals over a 20-min period to determine initial rates of repair. Strand-specific repair was scored as production of hydrolytic products upon cleavage with the marker restriction enzyme and a second endonuclease, which was *Cla*I in all cases except for the A-A (C^+/V^-) and T-T (C^-/V^+) heteroduplexes where secondary cleavage was accomplished with *Hin*cII. Rates of repair are expressed as femtomoles min^{-1} mg^{-1}. As discussed in the text and shown in Table III, repair events indicated with an * were found to be independent of the *mutS* gene product.

Heteroduplex	Phages	Markers	Methylation state			
			C^+/V^-		C^-/V^+	
			Rate of repair	Bias	Rate of repair	Bias
C 5′ cTCGA **G** AGCTt V 3′ gAGCT **T** TCGAa	f1MR3 f1MR1	*Xho*I *Hin*dIII	10.0	9	7.0	>15
C 5′ cTCGA **A** AGCTt V 3′ gAGCT **C** TCGAa	f1MR1 f1MR3	*Hin*dIII *Xho*I	6.1	>15	3.0	>15
C 5′ gTCGA **C** AGCTt V 3′ cAGCT **T** TCGAa	f1MR7 f1MR6	*Sal*I *Hin*dIII	6.7	7	3.9	3.5
C 5′ gTCGA **A** AGCTt V 3′ cAGCT **G** TCGAa	f1MR6 f1MR7	*Hin*dIII *Sal*I	6.1	1	9.2*	8
C 5′ aTCGA **A** AGCTt V 3′ tAGCT **A** TCGAa	f1MR4 f1MR5	*Hin*dIII *Cla*I	8.1	>15	4.4	7
C 5′ aTCGA **T** AGCTt V 3′ tAGCT **T** TCGAa	f1MR5 f1MR4	*Cla*I *Hin*dIII	7.9	8	3.6	6
C 5′ cTCGA **G** AGCTg V 3′ gAGCT **G** TCGAc	f1MR8 f1MR9	*Xho*I *Pvu*II	8.0	>15	3.3	>15
C 5′ cTCGA **C** AGCTg V 3′ gAGCT **C** TCGAc	f1MR9 f1MR8	*Pvu*II *Xho*I	0.2*	1	1.4*	3

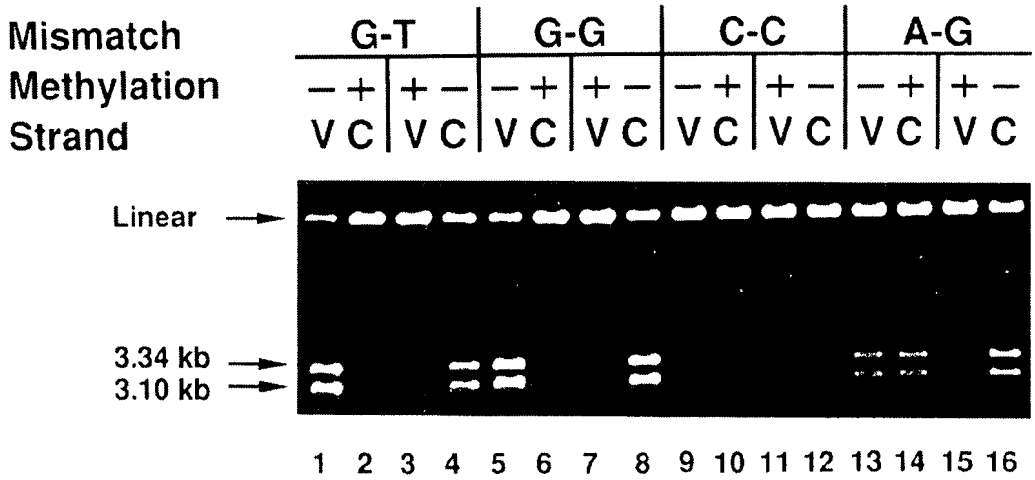


FIG. 2. **Well repaired, poorly repaired, and aberrantly repaired mismatches as revealed by *in vitro* assay.** Heteroduplex DNAs (10 μg/ml) as indicated were incubated at 37 °C for 1 h with extract prepared from *E. coli* AB1157 at a protein concentration of 10 mg/ml as described under "Experimental Procedures." After phenol extraction and ethanol precipitation (7), DNA samples were hydrolyzed with *ClaI* (position 2527) and the appropriate restriction endonuclease to monitor mismatch repair (Table II), and restriction products separated by electrophoresis in 1% agarose gels. *Arrows* indicate 3.34- and 3.10-kilobase fragments resulting from corrected heteroduplexes. *Lanes* marked *C* and *V* show repair products resulting from correction on the complementary or viral DNA strand, respectively. + and − indicate whether the particular DNA strand was methylated or unmethylated at the d(GATC) site at position 216.

TABLE III

The requirement for mutS gene product in the correction of mismatches

Mismatch correction of hemimethylated heteroduplex DNAs in *mutS*::Tn*5* (RK1517) cell extract was determined at a protein concentration of 10 mg/ml at 37 °C for 1 h as described under "Experimental Procedures" and Table II. For the complementation experiments, 50 ng of MutS protein was premixed with RK1517 cell extract (0.1 mg of protein) on ice, and repair reactions were initiated by adding the protein mixture to substrates preincubated at 37 °C. With the exception of the A-G mismatch, values shown are extents of repair on the unmethylated DNA strand after 1 h incubation. In the case of the A-G mispair, correction was determined on both DNA strands, with repair on the methylated DNA strand shown in the entry marked with an *. DNA strands and their state of methylation are designated as in Table II. NS, not significant.

Mismatch C-V	Repair C⁺/V⁻ MutS protein No	MutS protein Yes	MutS dependence	Repair C⁻/V⁺ MutS protein No	MutS protein Yes	MutS dependence
	fmol/mg		%	*fmol/mg*		%
G-T	<12	137	>90	12	98	85
A-C	17	120	86	19	103	81
A-A	<12	70	>83	<12	58	>79
G-G	<12	96	>88	19	74	75
T-T	<12	96	>88	<12	55	>78
C-T	<12	62	>81	17	74	77
C-C	<12	12	NS	12	14	NS
A-G	<12	36	>67	125	146	15
A-G*	103	115	10	<12	<12	NS

extracts. Although not shown, repair of these six mispairs also required the presence of *mutH* and *mutL* gene products.

In contrast to the behavior of the six mispairs described above, the C-C mismatch was an extremely poor substrate for mismatch correction (Fig. 2, *lanes 9–12*). As shown in Table II, this mispair was rectified at a rate of only 2–20% of that observed for the G-T mismatch, depending on the strand on which repair occurred. Furthermore, the low level repair of C-C displayed little bias for the unmethylated DNA strand (Table II), and this reaction was independent of the *mutS* (Table III), *mutH*, and *mutL* gene products (not shown).

Aberrant Repair of A-G Mismatch—The *in vitro* behavior of the A-G mismatch also differed from that observed for G-T, A-C, C-T, A-A, T-T, or G-G. As shown in Fig. 2 (*lanes 13* and *14*) and Table II, the A-G heteroduplex that was methylated on the complementary strand (the strand containing the A of the A-G mispair) was repaired on the methylated strand to yield a C-G pair at a rate comparable to that for correction of the unmethylated strand to yield an A-T base pair. The alternate heteroduplex bearing the methyl group on the viral strand (containing the G of the A-G mismatch) was subject to efficient, strand-specific correction to yield a C-G base pair.

These results reflect the operation of two distinct repair systems on the A-G mismatch *in vitro*. Correction of A-G to A-T in the heteroduplex methylated on the complementary DNA strand required the MutS protein (Table III), as well as *mutH* and *mutL* gene products (not shown). However, repair of A-G to C-G in either heteroduplex configuration was independent of the *mutS* gene product (Table III) and also did not require MutH or MutL proteins (not shown). Repair of the A-G heteroduplex methylated on the complementary strand therefore occurs by two distinct pathways. Correction on the unmethylated strand to yield an A-T pair is mediated by the methyl-directed, *mutHLS*-dependent pathway while repair of the methylated strand occurs by an alternate, *mut*-independent system which does not respond to d(GATC) methylation. The presence of a methyl group on the viral strand blocks repair on this strand by the *mutHLS* system, but the alternate repair pathway functions on the complementary strand to yield a C-G pair. Thus, the alternate, *mut*-independent pathway preferentially corrects the A-G mismatch to yield a C-G base pair. This system displays high specificity for A-G since the other mismatches located at the same position and within the same sequence environment were refractory to this sort of repair (Table II).

To determine whether *mut*-independent processing of A-G to C-G occurs in other sequence environments, we have constructed alternate heteroduplexes containing an A-G mismatch within overlapping sites for *Bsm*I and *Eco*RI endonucleases. Results obtained with this set of A-G heteroduplexes, which are summarized in Table IV, were similar to those for the A-G substrates described above. Repair to yield A-T required MutS protein and only occurred if the viral strand bearing the G of the A-G mismatch possessed an unmethylated d(GATC) sequence at position 216. Heteroduplexes in which this d(GATC) site was methylated on the viral strand or molecules which lacked a d(GATC) sequence at this position were not detectably corrected to an A-T base pair. In contrast, repair of the A-G to yield C-G occurred on all the heteroduplexes in a reaction that was independent of *mutS* function. In fact, repair of A-G to C-G appeared to be somewhat enhanced when methyl-directed correction to A-T was blocked (Table IV), suggesting that the two repair systems may compete for the A-G mispair.

MutS Protein Can Recognize All Eight Single Base Pair Mismatches—We have previously reported that the *E. coli* MutS protein binds in a specific manner to DNA fragments containing G-T, A-C, A-G, or C-T mismatches (12). However, a direct comparison of these four mismatches was compromised by the fact that they were located in different sequence environments. We have therefore extended this analysis to include the eight mismatches constructed in this study (Table II). As shown in Fig. 3, DNase I footprinting (22) indicated that all eight of these mispairs can be recognized by MutS protein. Although MutS protein was without effect on the DNase I hydrolytic pattern of the control restriction fragment, which contained a G-C pair instead of a mismatch, the protein did protect about 20 base pairs from hydrolysis when a mispair was present. We were unable to determine a more precise estimate of the size of the protected region due to the presence of a DNase I resistant region within this family of restriction fragments. This region is indicated by a *vertical bar* in Fig. 3.

MutS Protein Recognizes Mismatches

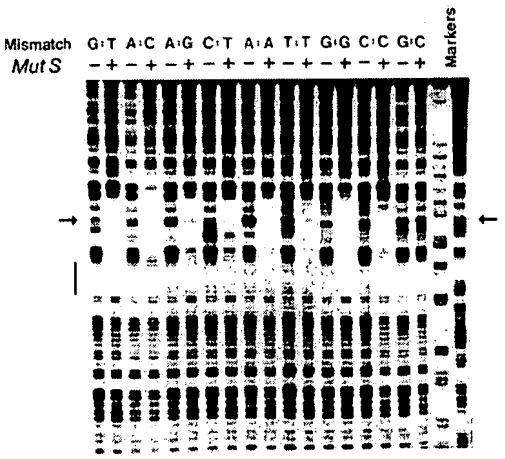


FIG. 3. **MutS protein protects regions containing a mismatch from DNase I hydrolysis.** A 143-base pair *Hae*II-*Ban*I fragment (coordinates 5572–5714), which was 3′-end-labeled at the *Ban*I terminus, was isolated from each of the eight f1MR heteroduplexes shown in Table II. The corresponding fragment from f1MR3 replicative form DNA served as control. This fragment contains a G-C pair at the location of the mismatch within the heteroduplexes (position 5632). DNase I footprinting reactions ("Experimental Procedures") contained 125 fmol of DNA fragment and either 3.0 pmol (G-T mismatch) or 4.0 pmol (other mismatches and G-C control fragment) of MutS protein. DNase I hydrolytic products were analyzed on an 8% polyacrylamide sequencing gel in parallel with markers prepared by chemical cleavage (G > A and T + C) reactions of Maxam and Gilbert (30). DNase I cleavage products derived from the 3′-end-labeled DNA were assumed to run one nucleotide slower than the corresponding marker fragments. *Plus* and *minus symbols* indicate the presence and absence of MutS protein. Mismatch location is shown by *arrows*, while the *vertical bar* indicates the DNase I-resistant region evident in the absence of MutS protein which is referred to in the text.

TABLE IV

In vitro correction of a second A-G mismatch in wild type and mutS extracts

Molecules containing an A-G mismatch at position 5620

C 5′ G A ATTC *Eco*RI′
V 3′ C G TAAG *Bsm*I′

within overlapping *Eco*RI and *Bsm*I recognition sites were prepared as described under "Experimental Procedures" using $f1h_1$ and $f1G18_1$ (one d(GATC) site heteroduplex) or $f1h_0$ and $f1G18_0$ (no d(GATC) site heteroduplex). The repair on complementary and viral DNA strands are shown in columns C and V, respectively. For heteroduplexes containing a d(GATC) sequence at position 216, the state of methylation of a DNA strand is indicated by + or −. The designation 0/0 indicates that the heteroduplex lacks this d(GATC) site. Heteroduplex repair ("Experimental Procedures") was determined in wild type (AB1157) or *mutS*::Tn*5* (RK1517) cell extracts as described in the legend to Table III. Values shown reflect extent of repair after 1 h incubation at 37 °C.

Methylation (C/V)	Repair					
	Wild type extract		*mutS*::Tn5 extract			
			No MutS protein		Add MutS protein	
	C	V	C	V	C	V
	fmol/mg					
+/+	36	<12	91	<12	92	<12
+/−	17	151	60	<12	65	55
−/+	77	<12	79	<12	84	<12
−/−	46	74	82	<12	79	50
0/0	34	<12	84	<12	91	<12

These experiments also demonstrated that the presence of a mismatch can alter the local sensitivity to DNase I hydrolysis in a manner that is dependent on the mispair. For example, the control restriction fragment and those containing G-T and A-C mispairs are identical (Table II) except for the presence of G-C, G-T, or A-C at the position indicated by the *arrow* in Fig. 3. Nevertheless, DNase I hydrolytic patterns obtained in the absence of MutS protein varied in the vicinity of this position for the three DNAs. Similar differences in local hydrolytic patterns are evident in comparisons of A-G and C-T heteroduplexes, A-A and T-T heteroduplexes, and G-G and C-C heteroduplexes (Fig. 3, lanes without MutS protein). DNA fragments within each of these pairs were identical except for the nature of their mismatch, with differences between heteroduplex pairs being limited to the fifth position on either side of the mispair (Table II). Apparently DNase I responds to variation in local structure or helix dynamics (reviewed in Ref. 4) associated with particular mispairs.

Affinities of MutS Protein for the Eight Base-Base Mismatches—We have used DNase I footprinting to estimate affinities of MutS protein for the set of eight base pair mismatches shown in Table II. In this approach, which is illustrated in Fig. 4 for the fragment containing the G-T mismatch, the extent of specific MutS·DNA complex formation was determined by densitometry of footprint patterns. Data obtained in this manner was fit (23) to the function

$$[\text{MutS}]_B[\text{DNA}]_T = [\text{MutS}]_F/(K + [\text{MutS}]_F)$$

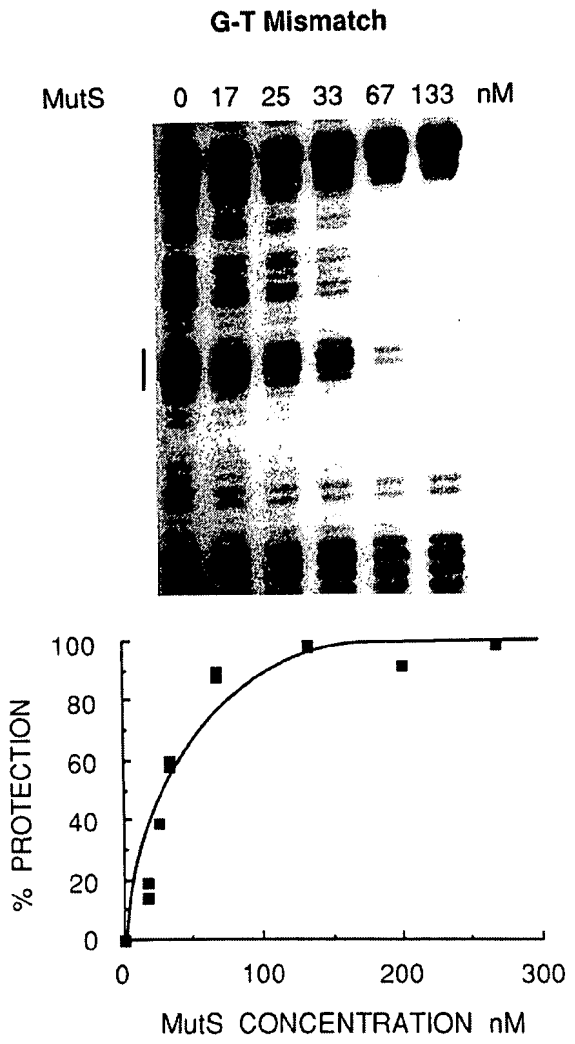


FIG. 4. **Estimation of binding affinity of MutS protein for a G-T heteroduplex.** DNase I footprinting reactions ("Experimental Procedures") contained a 3′-end-labeled *Hae*II-*Ban*I fragment containing the G-T mismatch at a concentration of 8.3 nM and MutS protein as indicated (concentration expressed as monomer equivalents). The extent of specific complex formation was estimated as the mole fraction of the DNA fragment which was protected against DNase I attack in the vicinity of the mismatch. This parameter was evaluated as the extent of protection of particular bands (*upper panel*) as determined by densitometry, as illustrated here for the G-T mismatch. The integrated intensity corresponding autoradiographic bands bracketed by the *vertical bar* was determined by scanning densitometry using a Zeineh Soft Laser Densitometer. To correct for variations in gel loading, the integrated intensity determined in this manner was normalized to that of a band lying outside the protected region (not visible on the photograph shown here). Percent protection shown in the *lower panel* was calculated as $100[1 - (I/I_0)]$ where I is the integrated intensity within the protected region at a given concentration of MutS protein, and I_0 is the corresponding intensity in the absence of added protein. The curve shown is that determined by Marquardt fit (23) of the function described in the text.

where $[MutS]_B$ is the concentration of specific MutS DNA complex, $[DNA]_T$ is the total concentration of DNA fragment, $[MutS]_F$ is the concentration of free MutS protein, and K is the equilibrium dissociation constant. The quotient to the left of the equal sign is assumed to be equivalent to the mole fraction of the DNA fragment protected against DNase I attack and $[MutS]_F$ to be equal to the total MutS concentration less that bound at the mismatch. Since we have not established the fraction of our MutS preparation that is biologically active nor the functional aggregation state of the protein, these calculations assume a monomer functional unit and that 100% of the protein is active in mismatch recognition. Binding constants determined in this manner may be in error due to these assumptions, but this approach should provide a reasonable picture of the relative affinities of the protein for the different mispairs.

TABLE V

Apparent affinities of mutS protein for base pair mismatches

The degree of MutS protection for each mismatch was analyzed and apparent dissociation constants evaluated as described in the text and the legend to Fig. 4.

Mismatch	Apparent dissociation constant
	nM
G-T	39 ± 4
A-C	53 ± 4
A-A	110 ± 7
T-T	140 ± 9
G-G	150 ± 10
A-G	270 ± 30
C-T	370 ± 40
C-C	480 ± 50

As summarized in Table V, apparent equilibrium dissociation constants for specific MutS complexes fell into three major classes. The protein displayed highest affinity for the G-T and A-C transition mismatches. A-A, T-T, and G-G were bound with intermediate affinity, and low affinity was observed for A-G, C-T, and C-C. These classes correlate reasonably well with the efficiencies of *in vitro* correction (Table II) with the exception of the A-C mismatch. Although initial rates of repair of the A-C and C-T mismatches were almost the same, the affinity of MutS for the two mismatches differed by 7-fold. As discussed below, such differences suggest that other interactions may also function in determining overall repair efficiencies.

DISCUSSION

The *in vitro* experiments described here indicate that the methyl-directed mismatch correction system of *E. coli* is capable of recognizing not only transition mismatches but also most of the transversion mispairs. This finding is consistent with previous analyses of the spontaneous mutation spectra of *mutH*, *mutL*, and *mutS* strains (24–26). These studies have demonstrated that increases in transition, transversion, and single base deletion mutations are associated with the mutator phenotype. However, transversions constitute only 3–4% of the base substitution mutations arising in such strains (25, 26). This is in contrast to *mut*$^+$ strains where transitions and transversions occur with approximately equal frequency. Such observations imply more efficient correction of transition mismatches, and Schaaper and Dunn have estimated approximate correction factors for the *mutHLS* pathway of 500–600-fold for transition mispairs and 30–40-fold for transverison mismatches.[1] The *in vitro* correction assay used in the work described here would be incapable of resolving such differences.

Our *in vitro* results may also be compared with specificity of *in vivo* correction as deduced from transfection experiments using artificially constructed heteroduplexes of λ (16, 17) or

[1] These correction factors should be regarded as rough estimates since the frequency of base substitution mutations can vary dramatically from site to site, presumably reflecting sequence environment effects.

M13 (15) DNA. These experiments indicate that G-T, A-C, A-A, and G-G are usually well corrected while C-T, A-G, T-T, and C-C mismatches are generally weaker substrates. However, some C-T, A-G, and T-T mispairs are well repaired, with differences in correction efficiency among this class presumably reflecting sequence environment effects (15–17). Thus, the *in vitro* results described here can be viewed as consistent with heteroduplex transfection studies. In making such comparisons it should be kept in mind that while the *in vitro* assay measures mismatch correction, the outcome of transfection experiments is determined not only by the rate of repair, but also by the rate at which DNA replication leads to segregation of the nucleotides comprising the mismatch. It is also pertinent to note that the λ heteroduplex studies, which have provided most of the information on specificity, utilized DNAs from phage mutants which arose spontaneously *in vivo* (16). Since these mutations presumably arose via pairing errors which went uncorrected in the *E. coli* cell, heteroduplexes constructed using such mutants may be enriched for mispairs which are less sensitive to repair.

Using hemimethylated heteroduplexes of the type shown in Fig. 1, we have consistently found that substrates methylated on the complementary strand are more efficiently repaired by the *mutHLS* pathway than the alternate configuration in which the methyl group resides on the viral stand (Table II). If methyl-directed mismatch repair involves unidirectional search or excision steps (4), then these differences in repair efficiency may reflect the different distances (1 *versus* 5.4 kilobases) between the mismatch and the d(GATC) site as measured along the two strands (Fig. 1). An alternate possibility is based on the previous observation that incision at a d(GATC) sequence by the MutH-associated endonuclease can occur with different efficiencies on the two DNA strands (14). In the case of the particular d(GATC) site present in the heteroduplexes described here, it has been found that the MutH-associated activity attacks the viral strand twice as fast as the complementary strand (14). If it is assumed that MutH-mediated cleavage at d(GATC) sites occurs as part of the repair process and that cleavage by this activity is at least partially rate limiting for repair, then this differential endonuclease sensitivity could account for the heteroduplex preferences shown in Table II.

The *in vitro* behavior of the two A-G mispairs that we have examined indicate that this mispair can be recognized and processed by two distinct systems. The *mutHLS*-dependent, methyl-directed pathway was found to correct the A-G mispairs to yield an A-T base pair. We have also identified an alternate, and apparently specific system which processes A-G to C-G by a mechanism that is independent of *mutH*, *mutL*, and *mutS* gene products and which does not depend on the presence of d(GATC) sequences. The biological significance of this *mut*-independent pathway is not clear since, to our knowledge, comparable effects have not been observed in heteroduplex transfection of *E. coli*.

A particularly surprising finding is that the MutS protein is capable of recognizing all eight of the base pair mismatches examined in this study. The several base-base mismatches that have been examined by physical methods have all been shown to be capable of adopting an intrahelical conformation (reviewed in Ref. 4), and physical and biological arguments have led to the suggestion that mismatches are recognized in this form (4, 27–29). However, the limited information available on the nature of mismatch structure and the lack of information on the nature of MutS-DNA interaction preclude consideration of the mechanism(s) by which a single protein is capable of recognition of the set of different mispairs.

Although C-C was weakly recognized by the MutS protein (Table V), significant *mutHLS*-dependent repair of this transversion mismatch was not observed. We therefore tested the possibility that the concentration of MutS protein present in extracts might be limiting for repair of this mismatch. However, supplementation of extracts with purified MutS protein (12) at concentrations approximately 2–10 times that of the endogenous level was without effect on the efficiency of correction of C-C, G-G, or G-T (not shown). Supplementation of extracts with a combination of near homogenous MutH (14), MutL,[2] and MutS (12) proteins resulted in only a small enhancement of G-T repair and was without significant effect on repair of G-G or C-C (not shown). Other arguments also indicate that mismatch recognition by MutS protein is not the sole factor in determination of correction efficiency. As discussed above, the rates of repair of the A-C and C-T mismatches are similar even though the affinity of MutS for A-C is substantially higher than that for the C-T mispair. The finding that hemimethylated heteroduplexes modified on the complementary DNA strand are more efficiently corrected than the alternate configuration also indicates that interactions other than mismatch recognition contribute to determination of repair efficiencies. The anomalous behavior of the two A-G mispairs described above is also consistent with this view.

REFERENCES

1. Claverys, J.-P., and Lacks, S. A. (1986) *Microbiol. Rev.* **50**, 133–165
2. Radman, M., and Wagner, R. (1986) *Annu. Rev. Genet.* **20**, 523–538
3. Meselson, M. (1988) in *The Recombination of Genetic Material* (Low, K. B., ed) Academic Press, Orlando, FL, in press
4. Modrich, P. (1987) *Annu. Rev. Biochem.* **56**, 435–466
5. Radman, M., Wagner, R. E., Glickman, B. W., and Meselson, M. (1980) in *Progress in Environmental Mutagenesis* (Alacevic, M., ed) pp. 121–130, Elsevier/North Holland, Amsterdam
6. Pukkila, P. J., Peterson, J., Herman, G., Modrich, P., and Meselson, M. (1983) *Genetics* **104**, 571–582
7. Lu, A.-L., Clark, S., and Modrich, P. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 4639–4643
8. Lahue, R. S., Su, S.-S., and Modrich, P. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 1482–1486
9. Nevers, P., and Spatz, H.-C. (1975) *Mol. Gen. Genet.* **139**, 233–243
10. Bauer, J., Krämmer, G., and Knippers, R. (1981) *Mol. Gen. Genet.* **181**, 541–547
11. Lu, A.-L., Welsh, K., Clark, S., Su, S.-S., and Modrich, P. (1984) *Cold Spring Harbor Symp. Quant. Biol.* **49**, 589–596
12. Su, S.-S, and Modrich, P. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 5057–5061
13. Lahue, R. S., Su, S.-S., Welsh, K., and Modrich, P. (1987) *UCLA Symp. Mol. Cell. Biol.* **47**, 125–134
14. Welsh, K. M., Lu, A.-L., and Modrich, P. (1987) *J. Biol. Chem.* **262**, 15624–15629
15. Kramer, B., Kramer, W., and Fritz, H.-J. (1984) *Cell* **38**, 879–887
16. Dohet, C., Wagner, R., and Radman, M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 503–505
17. Jones, M., Wagner, R., and Radman, M. (1987) *Genetics* **115**, 605–610
18. Marinus, M.G., Poteete, A., and Arraj, J. A. (1984) *Gene (Amst.)* **28**, 123–125
19. Benton, W. D., and Davis, R. W. (1977) *Science* **196**, 180–182
20. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
21. Bocke, J. D. (1981) *Mol. Gen. Genet.* **181**, 288–291
22. Galas, D. J., and Schmitz, A. (1978) *Nucleic Acids Res.* **5**, 3157–3170
23. Marquardt, D. W. (1963) *J. Soc. Ind. Appl. Math.* **11**, 431–441
24. Choy, H. E., and Fowler, R. G. (1985) *Mutat. Res.* **142**, 93–97
25. Leong, P.-M., Hsia, H. C., and Miller, J. H. (1986) *J. Bacteriol.* **168**, 412–416
26. Schaaper, R. M., and Dunn, R. L. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6220–6224
27. Kennard, O. (1985) *J. Biomol. Struct. Dyn.* **3**, 205–226
28. Hunter, W. N., Brown, T., Anand, N. N., and Kennard, O. (1986) *Nature* **320**, 552–555
29. Werntges, H., Steger, G., Riesner, D., and Fritz, H.-J. (1986) *Nucleic Acids Res.* **14**, 3773–3790
30. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560

[2] M. Grilley, K. Welsh, and P. Modrich, manuscript in preparation.

Volume 14 Number 15 1986 Nucleic Acids Research

Simultaneous synthesis of human-, mouse- and chimeric epidermal growth factor genes via 'hybrid gene synthesis' approach

Wing L.Sung*, Diana M.Zahab, Fei-L.Yao, Ray Wu+ and Saran A.Narang

Division of Biological Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6 and +Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received 3 March 1986; Revised and Accepted 30 June 1986

ABSTRACT

Simultaneous synthesis of two DNA duplexes encoding human and mouse epidermal growth factors (EGF) was accomplished in a single step. A 174 b.p. DNA heteroduplex, with 16 single and double base pair mismatches, was designed. One strand encoded the human EGF, and the opposite strand indirectly encoded the mouse EGF. The heteroduplex DNA was synthesized by ligation of seven overlapping oligodeoxyribonucleotides with a linearized plasmid. After transformation in E. coli HB101 (recA13), the resulting heteroduplex plasmid served as the template in plasmid replication. Two different plasmid progenies bearing either the human or mouse EGF-coding sequence were identified by colony hybridization using the appropriate probes. However, in E. coli JM103, the same process yielded plasmid progenies encoding different chimeric EGF molecules, presumably due to crossover of human and mouse EGF gene sequences.

INTRODUCTION

Modification of a gene in a controlled fashion is an important tool in the fields of molecular genetics and protein engineering. Small alterations can be obtained by either an oligonucleotide-directed "site-specific" mutagenesis or localized random mutagenesis of a gene already cloned in a plasmid vector (1-5). However, if extensive or multiple alterations are required, for example, in the preparation of homologous proteins of different species, the oligonucleotide-directed mutagenesis would appear to be too cumbersome. In this situation, separate synthesis of individual genes would be a more appropriate approach, albeit a laborious one (6).

We have recently devised a novel synthetic strategy, 'hybrid gene synthesis' (7), which produces both wild type and mutant genes simultaneously. The basic concept of this approach is to ligate several overlapping synthetic oligonucleotides

containing specific regions of mismatched bases to a linearized plasmid, yielding a heteroduplex plasmid. These mismatched regions are designed to generate both wild type and mutant genes. On transformation of a bacterial host with this heteroduplex plasmid, each plasmid strand will act as a template yielding two plasmid progenies bearing two related genes (1).

In this report, we have extended the hybrid gene synthesis approach to the simultaneous synthesis of sequences encoding human and mouse epidermal growth factors (hEGF and mEGF) (8,9). The two homologous polypeptides differ in sixteen amino acids out of fifty-three, amounting to a difference of 30% in the amino acid sequence (Figure 1).

By using a different bacterial strain in the transformation step, chimeric EGF genes were also generated. Our successful assembly of the coding sequences for both natural homologues of hEGF and mEGF, as well as the chimeric EGF, demonstrated the power of the hybrid gene synthesis approach.

METHODS AND MATERIALS

Enzymes and plasmid pUC8 were purchased from Bethesda Research Laboratories and Bcehringer Mannheim. Bacterial strains *E. coli* JM103 [Δ*lac pro*, *thi*, *straA*, *supE*, *endA*, *sbcB*, *hsdR*, F *tra* D36, *proAB*, *lac* I^q, ZΔM15] and *E. coli* HB101 [F^-, *hsd S* 20, *recA* 13, *ara*-14, *pro* A2, *lac* Y1, *gal* K2, *rps* L20, *xyl*-5, *mtl*-1, *sup* E44, λ^-] were used in the transformation experiments.

Synthesis of oligonucleotides

Seven oligodeoxyribonucleotides: EI, EII, EIIIa, EIIIb, EIV, EV and EVI (Figure 1) were synthesized by DNA synthesizer (model 380A Applied Biosystem). After deblocking, the synthetic fragments were purified on a 12% polyacrylamide gel containing 7M urea. These fragments constitute a heteroduplex which encodes hEGF with the frequently used yeast codons in one strand and mEGF as the complementary sequence in the opposite strand.

Construction of the plasmid heteroduplex bearing EGF sequences

Seven synthetic oligonucleotides EI, EII, EIIIa, EIIIb, EIV, EV and EVI (1 pmole, 1 μl) were phosphorylated individually in a final volume of 6 μl containing 0.6 μl of 10X kinase buffer, 0.6 μl of 1 mM ATP, 0.6 μl (6U) of T_4 polynucleotide kinase at

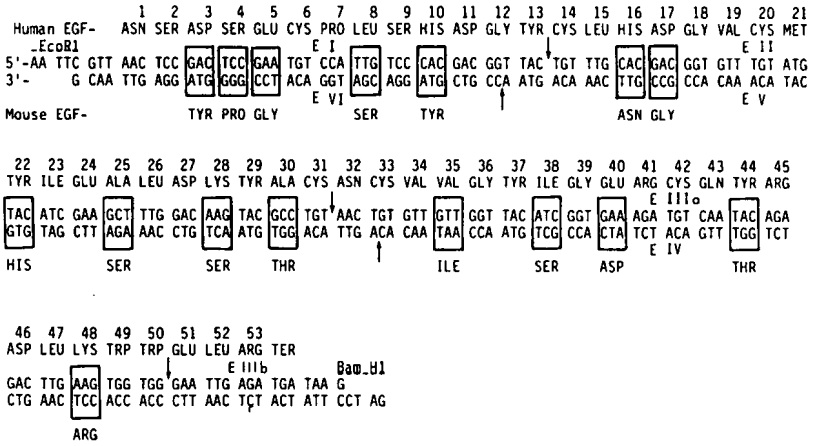


Figure 1. DNA heteroduplex encoding both human and mouse EGF. Upper strand encodes human EGF, and the lower strand encodes the mouse EGF as complementary sequence. Whenever amino acid residues in both h and mEGF differ, the triplet codons at those positions are contained in boxes. Polypeptide sequence of mouse EGF is not shown unless the amino acid residue is specific for mEGF. Ends of synthetic oligonucleotides constituting the DNA heteroduplex were indicated by arrows.

37° for 2 hr. These phosphorylation solutions were combined and heated at 70° for 10 min. After slow cooling to room temperature, the combined solution was added to a mixture of 0.1 pmole (2 µl) of EcoR1-BamH1 linearized plasmid vector pUC8, 6 µl of 4 mM ATP, 3 µl of 10X kinase buffer, 4 µl (12U) of T_4-DNA ligase and 15 µl of water. After incubation at 12° for 20 hr, aliquots of the ligation mixture were used to transform E. coli strains JM103 (X-gal, IPTG) and HB101 on YT agar plates containing ampicillin.

Screening of plasmid pxEGF bearing chimeric EGF-coding sequences

JM103 transformants which were white in color, resulting from a loss of β-galactosidase activity, were selected for colony hybridization analysis. Cells were grown overnight on seven identical nitrocellulose filters on YT plates containing ampicillin. The cells were lysed and the DNA in them was denatured with 0.5 N NaOH/1.5 N NaCl (10 min) and neutralized with 0.5 N Tris HCl (pH 7.0)/1.5 N NaCl (10 min). After 2 hr at 80° in vacuum oven, filters were washed with 6X SSC/0.05% Triton

X-100 for 30 min. The cell debris was scraped off and the filters were washed for another 30 min in fresh solution. The filters were then transferred individually into mixtures of 6X SSC/1% dextran sulfate/1X Denhardt hybridization fluid. Each of the seven ^{32}P-labelled probes EI-VI without purification were added to each filter. After 16 hr at 45°, the filters were washed twice with 6X SSC/0.05% Triton X-100 at room temperature for 5 min, and once at 45° for 45 min. The filters were analysed by autoradiography. Filters were washed again at 75° for 45 min for further autoradiographic analysis. Mini plasmid preparations were made on positive colonies. The EGF-coding region was sequenced by the dideoxynucleotide chain termination procedure.

Screening of plasmids pmEGF and phEGF bearing mEGF and hEGF-coding sequences

Colonies of transformed HB101 cells were chosen at random for hybridization analysis. The hybridization process was identical to the one used in the screening of plasmids pxEGF in the JM103 colonies. After washing at 75° for 45 min, colonies that hybridized to either E1-IIIa or EIV-VI were chosen for mini plasmid preparation and DNA sequencing.

RESULTS

Seven synthetic oligodeoxyribonucleotides, E1-IIIb encoding the human EGF, and EIV-VI encoding the mouse EGF as a complementary sequence, were phosphorylated and ligated to EcoR1-BamH1 cut plasmid vector pUC8 in a single operation (Figure 2).

Since the two strands of the resulting plasmid heteroduplex would serve as templates for plasmid replication, subsequent transformation of the bacterial host was expected to generate two different plasmid progenies bearing either the human or the mouse EGF gene.

The ligation mixture prepared was first used in the transformation of E. coli strain JM103 because of the convenience of using β-galactosidase activity as an indicator. Loss of this enzymatic activity in JM103 transformants immediately identifies those colonies that contain the EGF insert.

Colony hybridization with the two sets of probes, E1-IIIb and EIV-VI was used to distinguish transformants which

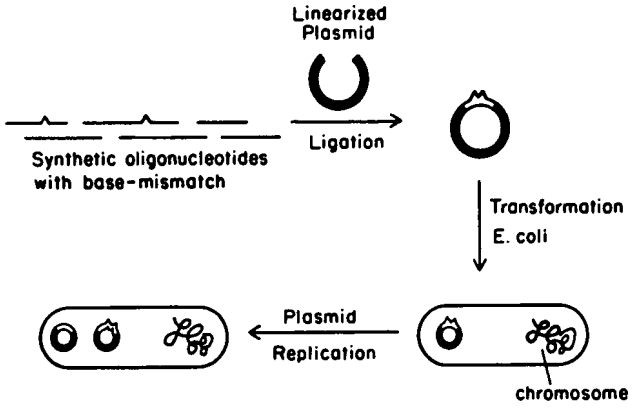


Figure 2. Strategy for the simultaneous synthesis of both human and mouse EGF coding sequences. Overlapping synthetic oligonucleotides with regions of base mismatch were phosphorylated, annealed and ligated to linearized plasmid vector to yield heteroduplex. Inside the bacterial host, the two strands of the heteroduplex plasmid would serve as templates for plasmid replication, and would yield two different progenies carrying either the mouse or human EGF coding sequence.

contained either the human or the mouse EGF gene. Each transformant selected for this analysis hybridized to either all or some of the labelled probes EIV-VI which constituted mouse EGF gene sequence.

The insert-bearing region of plasmid DNA from these transformants was sequenced by the dideoxynucleotide chain termination method. None of the twenty-four plasmid progenies chosen contained the correct coding seuqence of mouse or human EGF. Twelve plasmid progenies contained the general mouse EGF DNA sequences, but they also had many base deletions and substitutions. Among the remaining plasmid progenies, widespread interchange between the mouse and human EGF coding sequences has occurred to produce chimeric EGF genes (10) (Figure 3). In the three cases examined, the mouse EGF sequence constituted the major portion of the insert. Plasmid pxEGF-8 bore a hybrid EGF gene which encoded the first seven amino acids [1-7] of human EGF followed by forty-six amino acids [8-53] of mouse EGF (Figure 4). Plasmid pxEGF-16 encoded amino acids 1-15 of hEGF [1-15] and

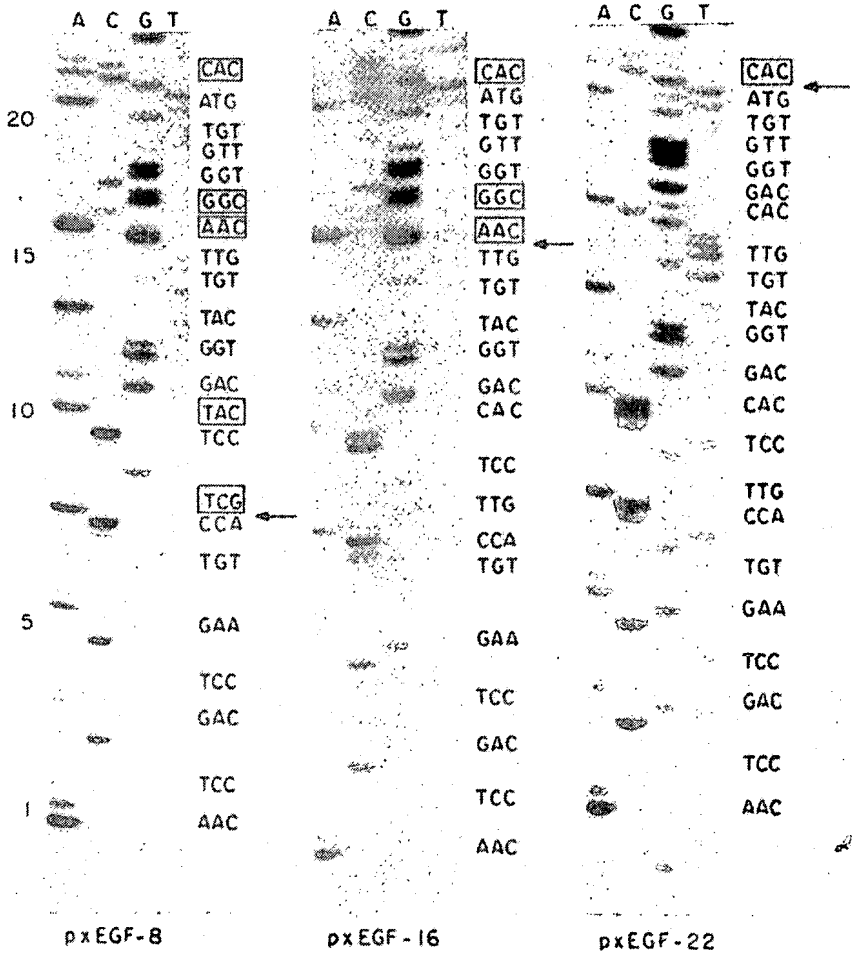


Figure 3. Autoradiogram of 5'-end of chimeric EGF coding sequences in plasmids pxEGF-22, pxEGF-16 and pxEGF-8. Triplet codons specific for mEGF are contained in boxes. Sequencing primer is 5'ACAGGAAACAGCTATGACC, annealed at the upstream region of the EcoRI cloning site. Numbers on left identify individual amino acid residues of EGF.

16-53 of mEGF. Another plasmid pxEGF-22 encoded amino acids 1-21 of hEGF and 22-53 of mEGF.

Because of widespread interchange of mouse and human EGF sequences in bacterial host JM103, the transformation experiment was repeated in another E. coli strain HB101 (recA 13) (11). In subsequent colony hybridization, four-fifths of the

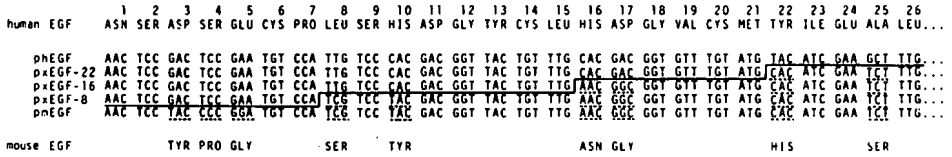


Figure 4. Nucleotide and encoded amino acid sequences of 5'-end of EGF gene in plasmids phEGF, pxEGF-22, pxEGF-16, pxEGF-8 and pmEGF. Triplet codons specific for mouse EGF were underlined by broken lines. Line in steps illustrates crossover of mouse and human EGF coding sequences in various plasmids containing chimeric EGF gene.

transformants hybridized to probes E1-IIIb, which encoded the human EGF, while the remaining transformants hybridized to probes EIV-VI which encoded the mouse EGF.

Eight progenies from both human and mouse series were then chosen for DNA sequence analysis of the EGF insert. DNA sequencing of plasmids phEGF and pmEGF confirmed that two progenies selected from each series contained the complete coding sequences of hEGF and mEGF, respectively (Figures 4 and 5). The hEGF or mEGF coding sequence in the remaining progenies was mutated by base deletions and substitutions. However, no interchange of mEGF and hEGF sequences could be detected in any plasmid progenies analysed.

DISCUSSION

All the gene assembly procedure reported so far involved the "one synthesis-one gene" strategy (12). One strand of the assembled gene corresponds to the message, the other strand is present only to maintain the duplex structure.

In the present study, we have extended the use of the hybrid gene synthesis approach (1) in preparing sequences encoding natural homologues of human and mouse EGF. By using this approach, a 174 b.p. DNA heteroduplex was designed to contain sixteen single and double base-mismatched regions. One strand of the heteroduplex directly encoded the human EGF and the other strand encoded mouse EGF in the complementary sequence (Figure 1). Transformation of a *recA* strain of bacterial host, *E. coli* HB101 (*recA* 13) successfully yielded plasmid progenies bearing either the human or mouse EGF gene.

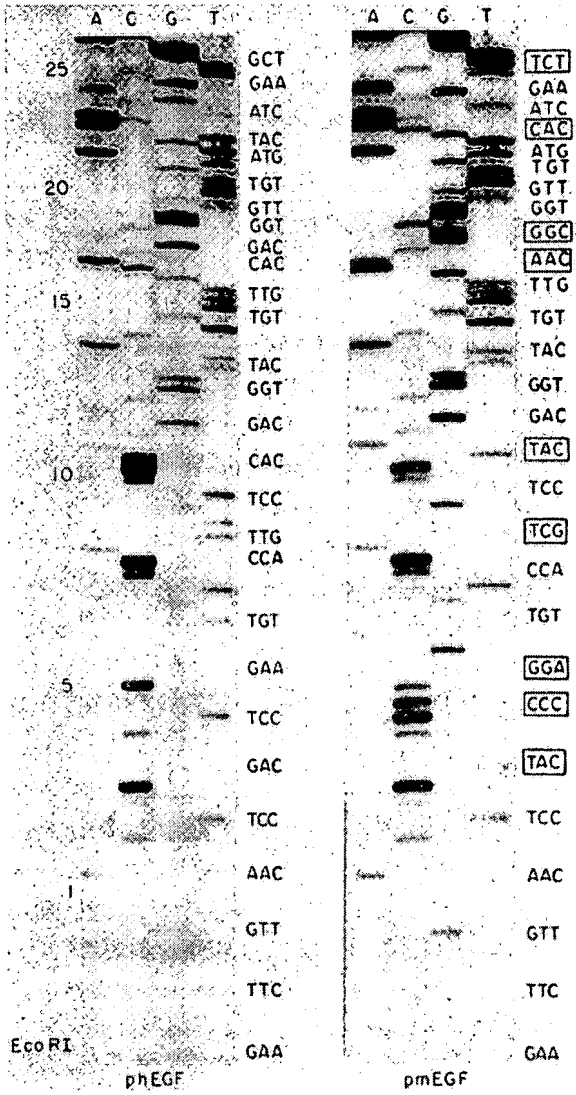


Figure 5. Autoradiogram of 5'-end of mouse and human EGF coding sequences in plasmids pmEGF and phEGF. Triplet codons specific for mEGF are contained in boxes. Sequencing primer is same as in Figure 3. Numbers on left identify individual amino acid residues of EGF.

When JM103 was used as host, plasmids containing mouse/human chimeric EGF sequences were found. This phenomenon was probably caused by the extensive base-mismatching regions of the plasmid heteroduplex and the use of a transformation host,

JM103, which favors DNA-repair or recombination (11,13).

Despite the exchange of genetic information in these chimeric genes, sequencing analysis indicated that the reading-frame of the DNA sequence was maintained in all cases examined. Therefore the expressed products of these chimeric genes would be genuine chimera or hybrids of natural homologous polypeptides.

Instead of starting from synthetic oligonucleotides, synthesis of chimeric genes has also been achieved in different families of interferon by either switching restriction fragments (14), or in vivo recombination between related genes already cloned in E. coli (15). In some cases, the chimeric products exhibited specificities different from either parent interferon molecules (16). Using our new strategy, it would be of great interest to see if more active growth factors, hormones or enzymes, in form of chimeric homologues could be generated. The present system of using a heteroduplex plasmid in a transformation host which favors DNA recombination would be a practical approach to generate a large number of chimeric molecules.

Further improvement of the hybrid gene synthesis, with its "one synthesis-multigenes" strategy to yield both homologous and chimeric gene sequences, would depend on a better understanding of the tolerance of base pair-mismatch in a duplex, and the in vivo DNA repair-recombination systems of the transformation host.

*To whom all the correspondence should be addressed. This is paper no. 25930 from the National Research Council of Canada.

REFERENCES

1. Gillam, S. and Smith, M. (1979) Gene 8, 99-106.
2. Miyada, C.G., Soberon, X., Itakura, K. and Wilcox, G. (1985) Gene 17, 167-177.
3. Morinaga, Y., Franceschini, T., Inouye, S. and Inouye, M. (1984) Bio/technology 2, 636-639.
4. Shortle, D., Koshland, D., Weinstock, G.M. and Botstein, D. (1980) Proc. Natl. Acad. Sci. USA 77, 5375-5379.
5. Botstein, D. and Shortle, D. (1985) Science 229, 1193-1201.
6. Jay, E., Rommens, J., Pomeroy-Cloney, L., Macknight, D., Lutze-Wallace, P., Wishart, P., Harrison, D., Lui, W.Y., Asundi, V., Dawood, M. and Jay, F. (1984) Proc. Natl. Acad. Sci. USA 81, 2290-2294.

7. Sung, W.L., Zahab, D.M., Yao, F.-L. and Tam, C.S. (1986) Biochem. Cell. Biol. 64, 133-138.
8. Sumi, S.-I., Hasegawa, A., Yagi, S., Miyoshi, K., Kanezawa, A., Nakagawa, S., Suzuki, M. (1985) J. Biotechnology 2, 59-74.
9. Savage, C.R. Jr., Inagami, T. and Cohen, S. (1972) J. Biol. Chem. 247, 7612-7621.
10. Radding, C.M. (1978) Ann. Rev. Biochem. 47, 847-880.
11. Clark, A.J. (1973) Ann. Rev. Biochem. 42, 67-86.
12. Crea, R., Hirose, T., Kraszewski, A. and Itakura, K. (1978) Proc. Natl. Acad. Sci. USA 75, 5765-5769.
13. Soltis, D.A. and Lehman, I.R. (1983) J. Biol. Chem. 258, 6073-6077.
14. Morrison, S.L. (1985) Science 229, 1202-1207.
15. Weber, H. and Weissmann, C. (1983) Nucleic Acids Research 11, 5661-5669.
16. Weck, P.K., Apperson, S., Stebbing, N., Gray, P.W., Leung, D., Shepard, H.M. and Yoeddel, D.V. (1981) Nucleic Acids Research 9, 6153-6165.

Volume 11 Number 16 1983 Nucleic Acids Research

Formation of genes coding for hybrid proteins by recombination between related, cloned genes in *E. coli*

Hans Weber and Charles Weissmann

Institut für Molekularbiologie I, Universität Zürich, 8093 Zürich, Switzerland

Received 4 July 1983; Accepted 19 July 1983

ABSTRACT

We describe a method for the formation of hybrid genes by in vivo recombination between two genes with partial sequence homology. DNA structures consisting of plasmid vector sequences, flanked by the α2 interferon gene on the one side and a portion of the α1 interferon gene (homology about 80%) on the other, were transfected into E.coli SK1592. Appropriate resistance markers allowed the isolation of colonies containing circular plasmids which arose by in vivo recombination between the partly homologous interferon gene sequences. Eleven different recombinant genes were identified, six of which encoded new hybrid interferons not easily accessible by recombinant DNA techniques.

INTRODUCTION

Twelve or more α-interferons are encoded in the human genome (for a review, see ref. 1), most of which are expressed to some degree (2,3,4). Some of these exhibit widely different antiviral activities on cultured cells of different animal origin (5,6,7). For example, the specific activity of interferon α1 is comparatively low on human cells but high on mouse cells, whereas the opposite is the case for interferon α2 (5,6,7). Genetic engineering techniques were used in vitro to construct hybrid interferon genes consisting of α1 and α2 specific sequences joined at either of two restriction sites present in both genes at homologous positions (5,6,7). The specific antiviral activities of hybrid interferons containing the C-terminal half of interferon α1 were high on mouse cells; on the other hand, high activity on human cells was dependent on the presence of the N-terminal half of interferon α2.

The types of recombinants that can be created by this approach are limited by the number of appropriate restriction

sites. As the nucleotide sequence homology between the genes for interferon α1 and α2 is about 80% it seemed likely that the recombination machinery of E.coli cells could be used in vivo to obtain recombinant genes with crossovers at any site showing sufficient homology.

In this paper we show that by an appropriate choice of constructions it is possible to generate with minimal effort a variety of hybrids not easily obtained by ordinary genetic engineering techniques. The positions of the desired crossovers can be directed to predetermined regions. The method should be generally applicable to the formation of recombinants between not too distantly related genes.

MATERIALS AND METHODS

1) Plasmids and bacteria.

Plasmids pM11 and pM21 (see description in Results and Discussion section) were constructed by Dr. M. Mishina (unpublished results). For the construction of plasmid pM11kan a derivative of plasmid pBR322 conferring resistance to tetracycline, ampicillin and kanamycin was first prepared. pBR322 was partially digested with HaeII in presence of 50 μg/ml ethidium bromide, to yield mostly full-size linear molecules. These were ligated with an equivalent amount of a complete HaeII digest of a pCRI plasmid (containing, for incidental reasons, a bacteriophage Qβ cDNA insert; kindly supplied by Dr. M. Billeter). Transfectants were selected for resistance against kanamycin and ampicillin, and subsequently tested for tetracycline resistance. For unknown reasons simultaneous selection for resistance against all three antibiotics yielded no clones. Restriction mapping showed that the 1430 bp HaeII fragment conferring kanamycin resistance had been inserted in pBR322 at the HaeII site in position 2352. Plasmid pM11kan was then constructed by ligating the following 3 components: (1) The 4400 bp PstI-SalI fragment of the kanamycin-resistant pBR322 derivative, (2) the 1000 bp ClaI-PstI fragment of pM11, (3) the ClaI-SalI fragment (about 800 bp) of a pBR322 derivative carrying a small insertion (a small lambda DNA fragment linked by poly(dG:dC) tracts) in the BamHl site. The 3 fragments were isolated by electrophoresis on 1% low-melting agarose gels

(SeaPlaque agarose, FMC Corporation). Aliquots of the gel slices were melted at 65° and used directly for the ligation reaction (8). Transfectant colonies were selected for kanamycin resistance and tested for tetracycline and ampicillin sensitivity. The structure of the desired plasmid pM11kan was confirmed by restriction mapping.

E.coli SK1592, a phage T_1-resistant derivative of strain SK1590 (9), was obtained from Dr. S. Kushner (University of Georgia) and E.coli 803 from Dr. K. Murray (University of Edinburgh). Enzymes were purchased from New England Biolabs Inc. and labeled nucleotides from Amersham.

2) In vivo recombination.

The SalI-PstI fragment (1580 bp) of pM21 and the SalI-BglII fragment (4970 bp) of pM11kan were isolated by electrophoresis on low melting agarose (1.2% and 0.8%, respectively) in 40 mM Tris-acetate, 1 mM EDTA (pH 7.8). Ligation was in reaction mixtures (20 µl) containing 20 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, 10-20 fmol of each DNA fragment (2-5 µl aliquots of agarose slices melted at 65°, final concentration of agarose 0.35%) and DNA ligase (400 units, New England Biolabs) at 16° for 18 h. Aliquots (2.5 µl) were used in transfection mixtures (100 µl) containing 10 mM $CaCl_2$, 10 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.5), and 65 µl of a suspension of $CaCl_2$-treated E.coli (prepared according to ref. 10 and stored in 50 mM $CaCl_2$, 20% glycerol at -80°). The suspensions were kept at 0° for 15 min, at 42° for 2 min, then diluted with L-broth (1 ml) and incubated with shaking at 37° for 90 min. Aliquots were spread on agar plates containing L-broth plus kanamycin (50 µg/ml) and tetracycline (20 µg/ml). The number of colonies varied between 1'600 and 90'000 per pmol DNA fragments; control plates with tetracycline alone gave $5\text{-}12\times10^7$ colonies/pmol pBR322. Plasmids were prepared by modifications of published methods (11,12). DNA sequences were determined as described by Guo and Wu (13), using the EcoRI, BglII or PvuII sites as first or second cleavage sites.

RESULTS AND DISCUSSION

The principle of the experiment is outlined in Fig. 1. The two parental interferon genes (or overlapping portions thereof)

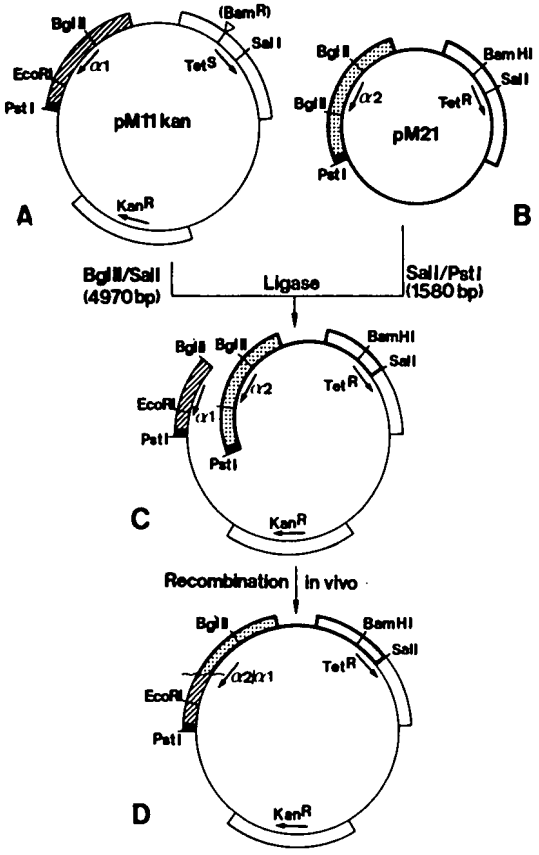
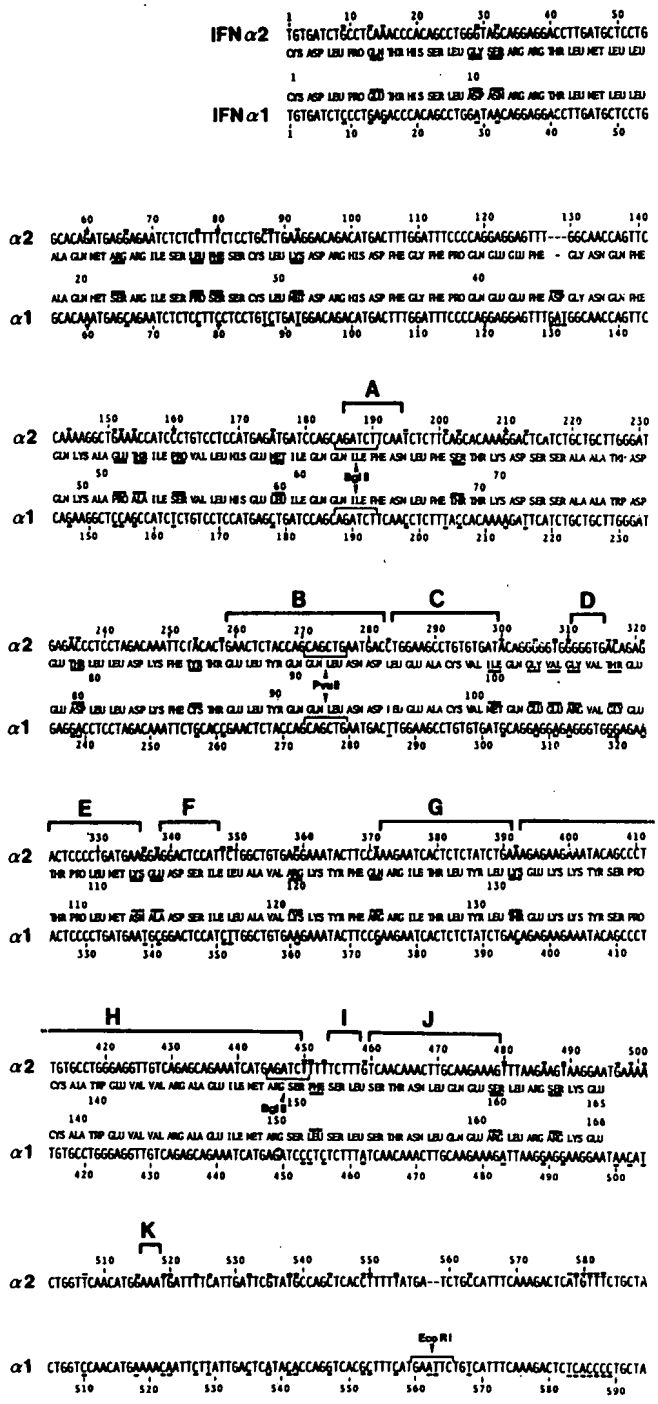


Figure 1. Construction of DNA amenable to recombination in vivo to yield α2-α1-interferon hybrid genes. A. Plasmid pM11kan is a pBR322-derived expression plasmid for interferon α1 production in E.coli. It contains a DNA segment conferring kanamycin resistance in position 2352 and has a small insertion in the BamH1 site, which abolishes tetracycline resistance. The mature interferon α1 coding sequence is joined to the β-lactamase promoter and ribosome binding region and replaces the N-terminal portion of the β-lactamase gene. B. In plasmid pM21 the interferon α2 gene is inserted in the same way into pBR322 as α1 in pM11kan; the remaining pBR322 sequences are unmodified. C. Appropriate fragments from both parent plasmids are ligated to yield linear (probably concatenated) DNA structures linked via the SalI site, thus reconstituting the tetracycline resistance gene. Details are given in the text. D. General structure of the interferon α2-α1 recombinant plasmids. The dotted regions indicate α2-specific, the hatched regions α1-specific sequences. The black regions adjacent to the PstI sites indicate short poly(dG:dC) tracts (15-30 bp). Bam^R, destroyed BamH1 site; Tet^R, Kan^R, tetracycline and kanamycin resistance genes; Tet^S, inactivated tetracycline resistance gene.

with the vector sequences between them are supplied to the host cell as parts of a linear DNA structure. Circularization of such structures by recombination within the interferon genes leads to replicating plasmids. An appropriate arrangement of two antibiotic resistance genes allows the easy selection of recombinants.

The linear DNA structures (presumably concatenates of the molecules shown in Fig. 1C) were constructed by ligation of two restriction fragments, each derived from a plasmid containing one of the parental interferon genes and an antibiotic resistance gene. Plasmid pM21, constructed for high expression of interferon α2 in E.coli (M. Mishina, unpublished results), contained the DNA sequence coding for mature interferon α2, fused to an AUG triplet 6 nucleotides downstream from the Shine-Dalgarno sequence of the β-lactamase gene of plasmid pBR322. The interferon segment replaced the β-lactamase sequences up to the PstI site, to which it was joined by a poly(dG:dC) tract. The remainder of the plasmid was identical to pBR322, including the intact tetracycline resistance region. In the other parental plasmid, pM11kan, the interferon α1 sequence was inserted into pBR322 in exactly the same way as α2 in pM21. In addition, the kanamycin resistance gene of plasmid pCRI (1430 bp HaeII fragment; ref. 14) had been inserted into the HaeII site at position 2352 and the tetracycline resistance region inactivated by a small insertion into the BamHI site.

To obtain linear DNA structures suitable for in vivo recombination between the α1 and α2 genes, the SalI-PstI fragment (1580 bp) of plasmid pM21 was ligated to the SalI-BglII fragment (4970 bp) of plasmid pM11kan and the resulting DNA transfected into $CaCl_2$-treated E.coli SK1592. Selection on agar plates containing both tetracycline and kanamycin yielded only bacteria transformed by ligation products containing elements of both parental moieties linked correctly at the SalI site, as the proximal portion of the tetracycline resistance region has to be provided by the α2 parent plasmid whereas the distal portion of the tetracycline resistance region as well as the kanamycin resistance gene originate from the α1 parent. As the IFN gene segments were the only homologous regions on the linear ligation products, it was expected that circularization by recombination


IFN α2
IFN α1
A
B
C
D
E
F
G
H
I
J
K
Pvu II
Bgl II
Eco RI

would occur predominantly if not exclusively between the BglII site (around position 190) and the end of the interferon DNA, i.e. the PstI site.

Plasmids from tetracycline and kanamycin resistant clones were subjected to restriction and sequence analysis. Of 63 clones analyzed, 62 appeared to have arisen by correct homologous recombination, i.e., without any gaps or insertions, as judged by restriction mapping and by the finding that 20 out of 20 cell-free extracts from the bacterial clones had levels of interferon activity similar to the α2 parent strain when tested on human HEp2 (CCL23) cells. Thirteen recombinants lacked the unique EcoRI site located downstream of the IFN-α1 coding sequence and probably contained entirely α2-specific coding sequences; they were not analyzed further. Forty-four of the remaining plasmids were sequenced between the upstream BglII site and the EcoRI site. As shown in Fig. 2, 11 different crossover regions were identified in this interval. The 11 recombinant sequences encoded 9 different interferons, 8 of which were hybrids. Only two of these had been obtained previously using conventional genetic engineering techniques (5,6,7).

Five of the clones yielded plasmid DNA preparations which were heterogeneous with regard to the downstream BglII site, one also with regard to the EcoRI site. Heterogeneity could be due to segregation from a heteroduplex recombination intermediate or to transfection by several copies of (possibly concatenated) recombination-competent DNA. Homogeneous plasmid preparations were obtained after retransfection and cloning, except for one subclone, where the BglII restriction pattern again indicated heterogeneity, with two components being present in about equal quantity. Since this DNA preparation consisted mainly of dimeric plasmids, it seemed likely that two different recombinants were linked in a tandem dimeric circle.

The number of recombinant plasmids recovered for each cross-

Figure 2. Location of 11 crossover regions (A through K) between the interferon α2 and α1 genes observed in the *in vivo* recombination experiments. Recombinants with crossovers in regions D, E, F, G, H and I/J code for new hybrid interferons. Crossovers in B and C as well as I and J give sequences differing only by silent nucleotide changes.

over region varied greatly, however, the experimental conditions used for selection did not ensure that all hybrids were independent isolates. The dependence of recombination frequency on the degree of homology, on the length of the homologous region and on specific sequence features remains to be determined. In any event, crossovers were found in regions with as few as five or even three bp of uninterrupted homology (Fig. 2, regions D and K). It should be noted however that the actual crossover point need not correspond to the region in which recombination is initiated, and that recombination might require regions of higher homology.

In additional experiments the actual recombination step between the plasmid components was carried out in vitro leaving only heteroduplex repair to take place in the host cell. Plasmid pM11kan was linearized at the BglII site (position 189); plasmid pM21 was partially digested with BglII to yield mostly linear full-size molecules, 60-70% of which were cut at the downstream BglII site (position 446) as shown by restriction analysis. Both linearized plasmid preparations were digested with T_4-DNA polymerase to convert the terminal 300-400 nucleotides into a single-stranded form; they were mixed in about equal proportion and annealed. Remaining gaps were filled in with DNA polymerase, the products were cleaved with restriction endonuclease SalI and circularized by ligase. Transfections were carried out into E.coli strains HB101 ($recA^-$), 803 (rec^+) and SK1592 ($recA^+sbcB$). Colonies resistant to both tetracycline and kanamycin were obtained from all three strains, but among 21 plasmids analyzed all but one, a hybrid generated in E.coli HB101 with a crossover in region C, were found by restriction mapping to contain deletions or otherwise rearranged sequences. It would seem that the in vivo process is both more efficient and simpler to carry out, but that in principle similar results could be produced in vitro.

The method we have described can be further refined to yield crossovers in predetermined regions. For example, if the linear concatemers are formed between the pM11kan SalI-BglII fragment and a pM21 fragment extending from the SalI to the PvuII site in position 273, recombination would be confined to

the 89 bp region between the BglII and the PvuII sites.

It will be of interest to determine how much homology between two genes is required to allow this type of recombination.

ACKNOWLEDGEMENTS

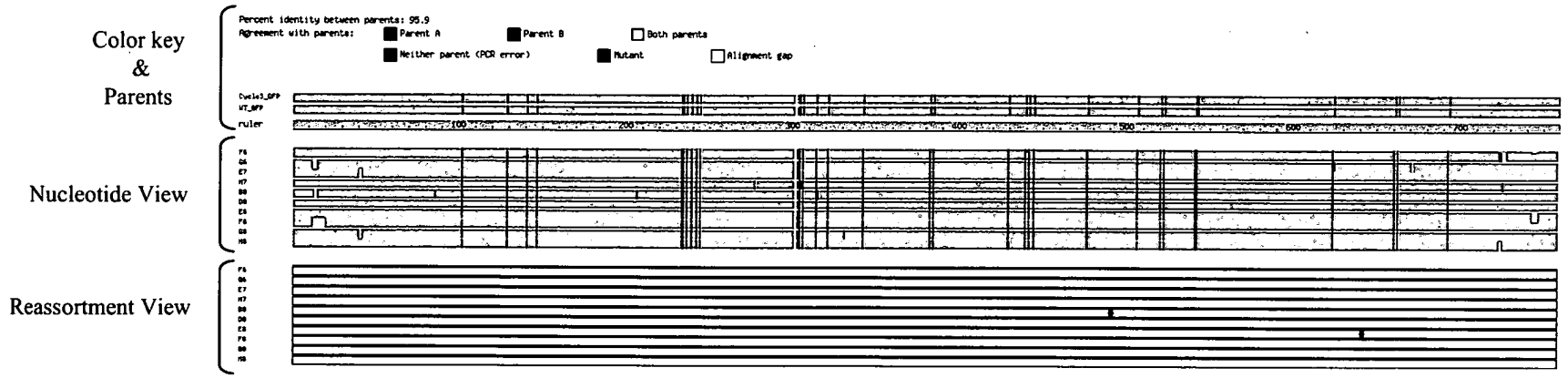
We thank Mr. Peter Meyer for competent technical assistance, Dr. David Valenzuela for analyzing several of the hybrid clones and Mr. Michel Streuli for $CaCl_2$-treated E.coli SK1592. This work was supported by the Schweizerische Nationalfonds and the Kanton Zürich.

REFERENCES

1. Weissmann, C., Nagata, S., Boll, W., Fountoulakis, M., Fujisawa, A., Fujisawa, J.-I., Haynes, J., Henco, K., Mantei, N., Ragg, H., Schein, C., Schmid, J., Shaw, G., Streuli, M., Taira, H., Todokoro, K. and Weidle, U. (1982) Phil. Trans. R. Soc. Lond. B299, 7-28.
2. Rubinstein, M., Levy, W.P., Moschera, J.A., Lai, C.-Y., Hershberg, R.D., Bartlett, R.T. and Pestka, S. (1981) Archs. Biochem. Biophys. 210, 307-318.
3. Allen, G. and Fantes, K.H. (1980) Nature 287, 408-411.
4. Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandliss, R., Seeburg, P.H., Ullrich, A., Yelverton, E. and Gray, P.W., (1981) Nature 290, 20-26.
5. Streuli, M., Hall, A., Boll, W., Stewart II, W.E., Nagata, S. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA 78, 2848-2852.
6. Weck, P.K., Apperson, S., Stebbing, N., Gray, P.W., Leung, D., Shepard, H.M. and Goeddel, D.V. (1981) Nucl. Acids Res. 9, 6153-6166.
7. Rehberg, E., Kelder, B., Hoal, E.G. and Pestka, S. (1982) J. Biol. Chem. 257, 11497-11502.
8. Frischauf, A.M., Garoff, H. and Lehrach, H. (1980) Nucl. Acids Res. 8, 5541-5549.
9. Kushner, S.R. (1978) In "Genetic Engineering" (H.W. Boyer and S. Nicosia, eds.) Elsevier North Holland, Amsterdam, pp. 17-23.
10. Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) Proc. Natl. Acad. Sci. USA 69, 2110-2114.
11. Currier, T.C. and Nester, E.W. (1976) Anal. Biochem. 76, 431-441.
12. Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193-197.
13. Guo, L.-H. and Wu, R. (1982) Nucl. Acids Res. 10, 2065-2084.
14. Oka, A., Sugisaki, H. and Takanami, M. (1981) J. Mol. Biol. 147, 217-226.

Figure A

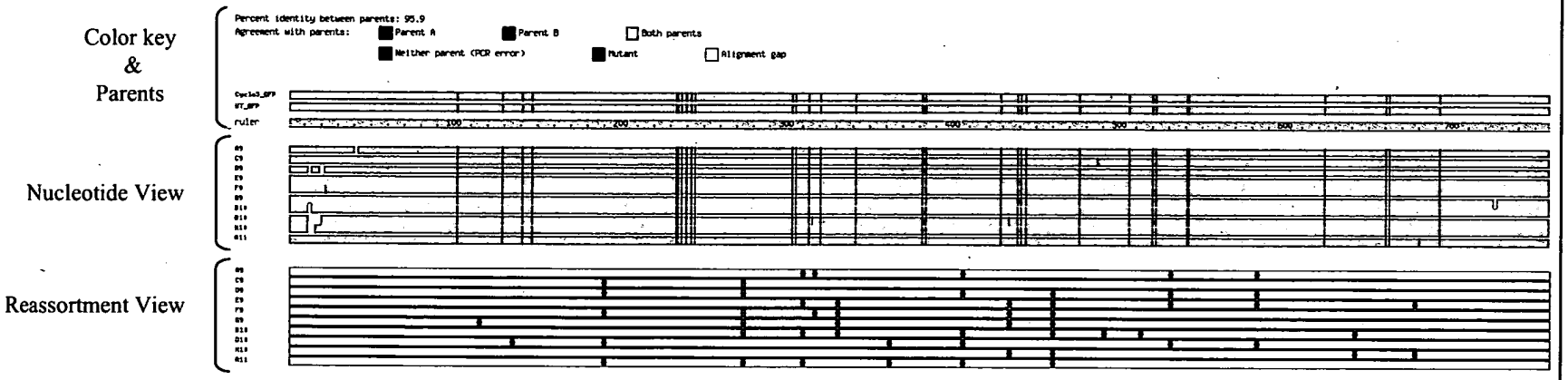
No GRAMMR Treatment - Negative Control



(CEL I omitted from GRAMMR reaction)

Two Crossovers in 10 Progeny

GRAMMR Treatment

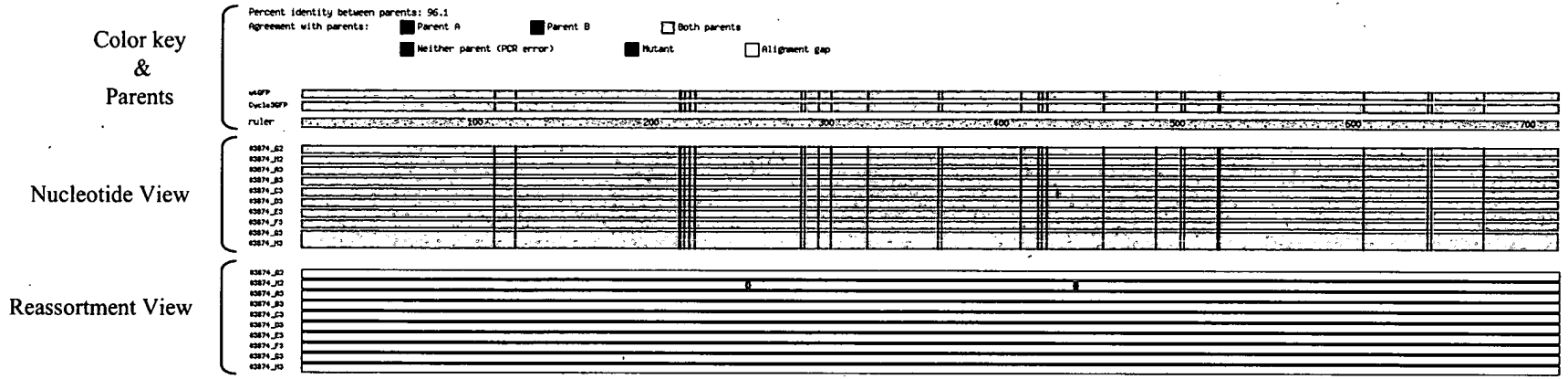


55 Crossovers in 10 Progeny

(5/gene; 9/kilobase)

Figure B

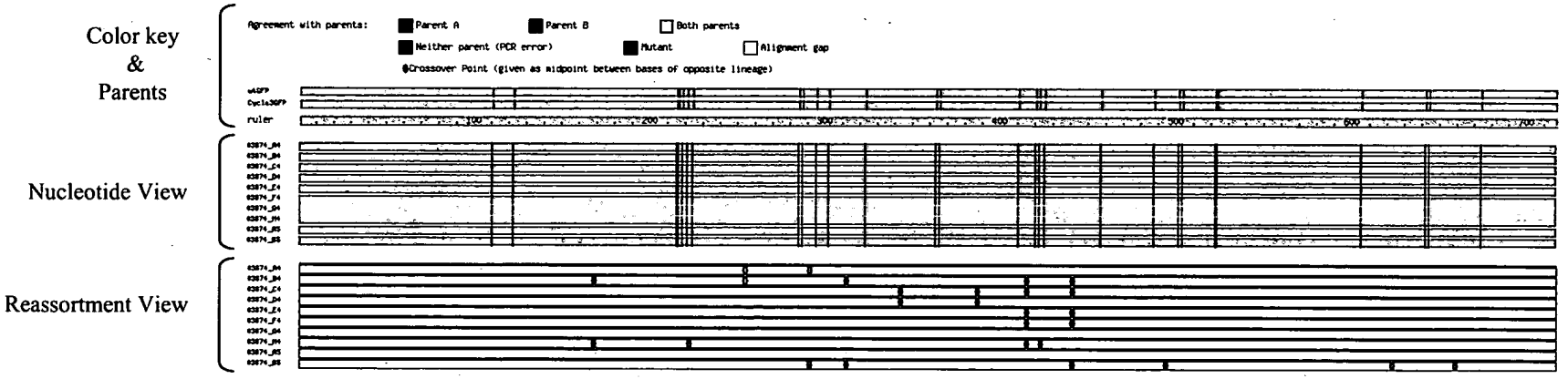
No GRAMMR Treatment - Negative Control



(CEL I omitted from GRAMMR reaction)

1 shuffled clone in 10 Progeny

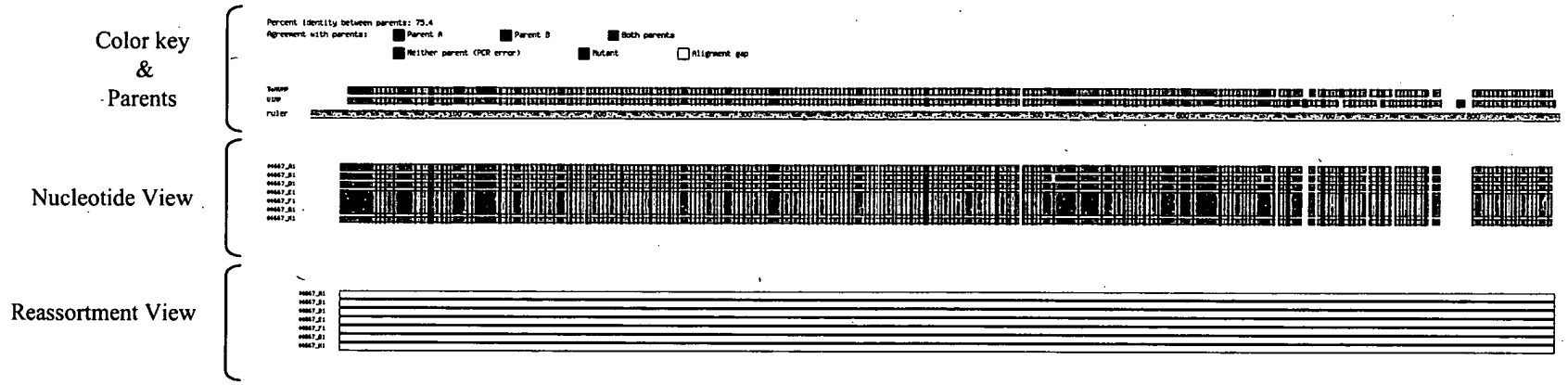
GRAMMR Treatment



27 Crossovers (2.7/gene; >4/kilobase)

Figure C

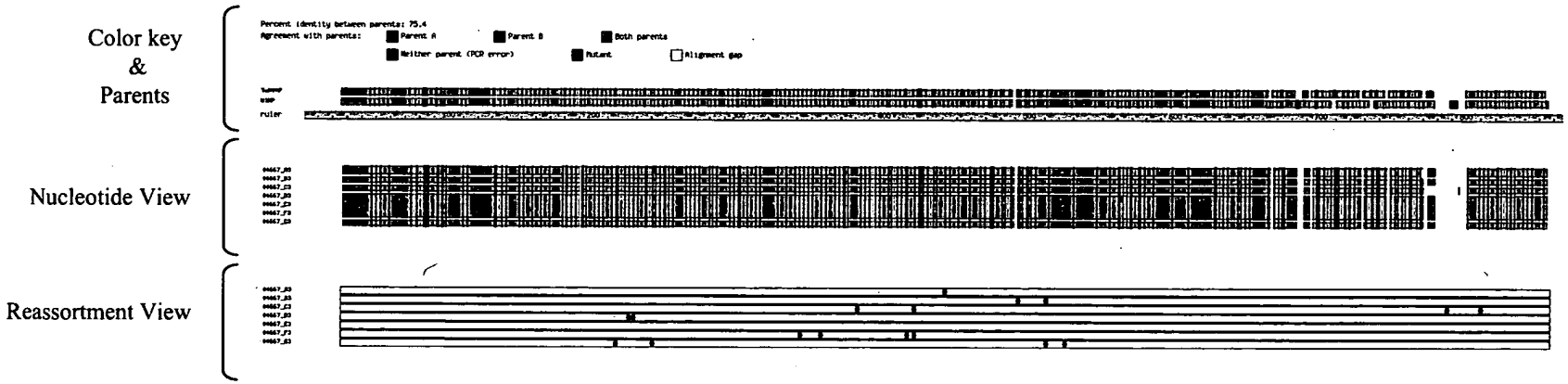
No GRAMMR Treatment - Negative Control



(CEL I omitted from GRAMMR reaction)

Zero Crossovers in 8 Progeny

GRAMMR Treatment

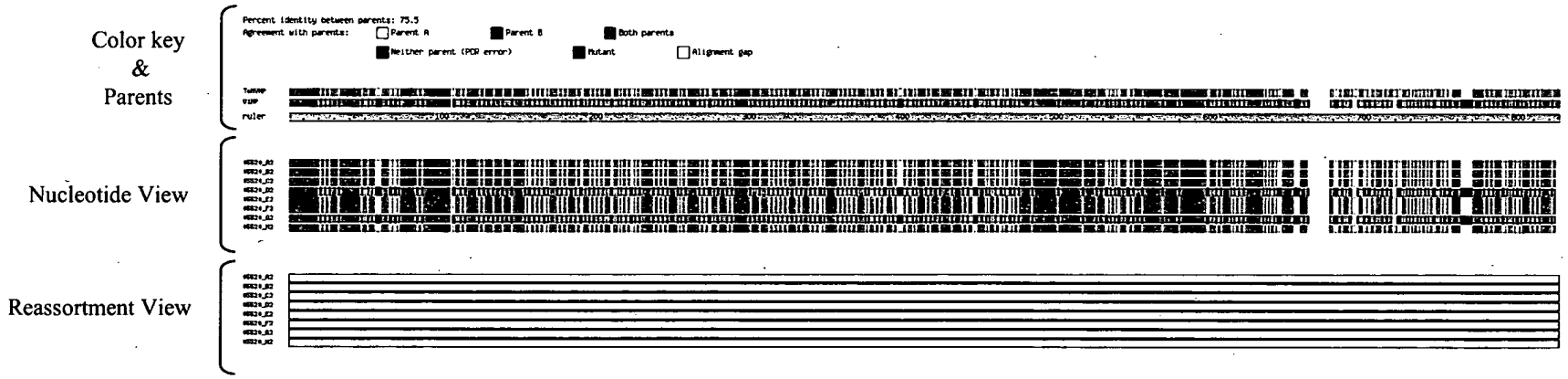


21 Crossovers in 8 Progeny

(>2.5/gene; >3/kilobase)

Figure D

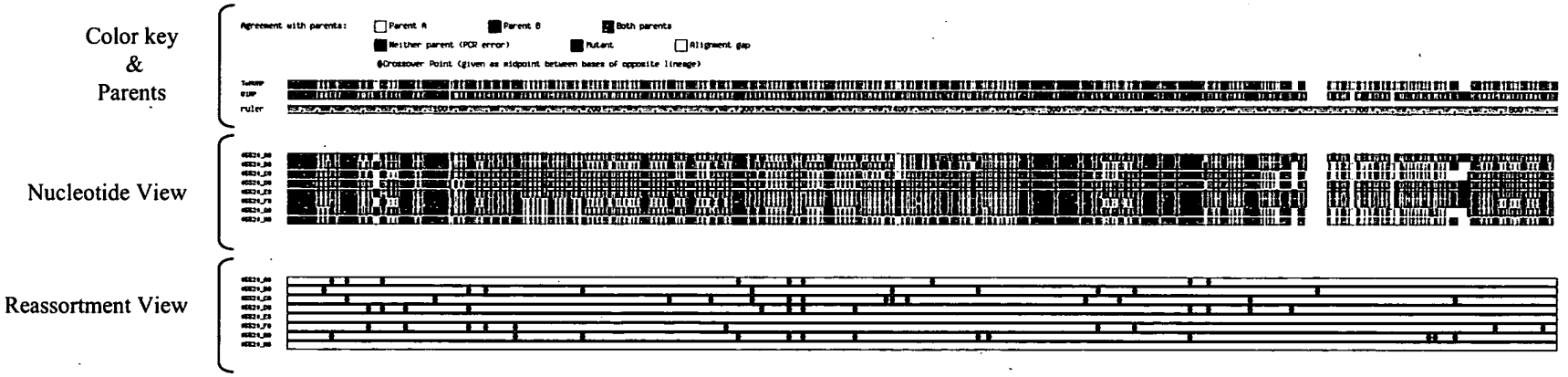
No GRAMMR Treatment - Negative Control



(CEL I omitted from GRAMMR reaction)

Zero Crossovers in 8 Progeny

GRAMMR Treatment

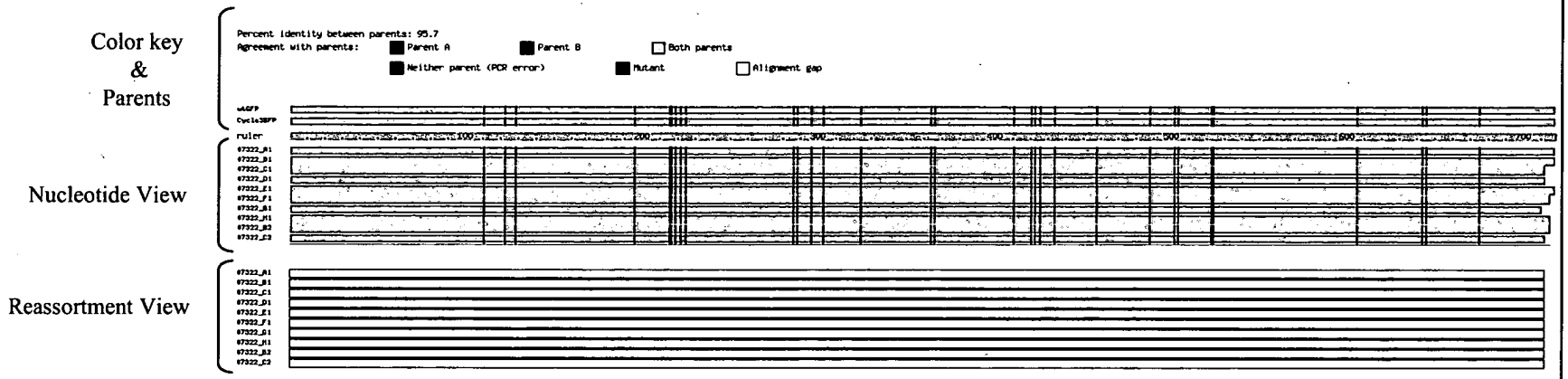


68 Crossovers in 8 Progeny

(>8/gene; >10/kilobase)

Figure E

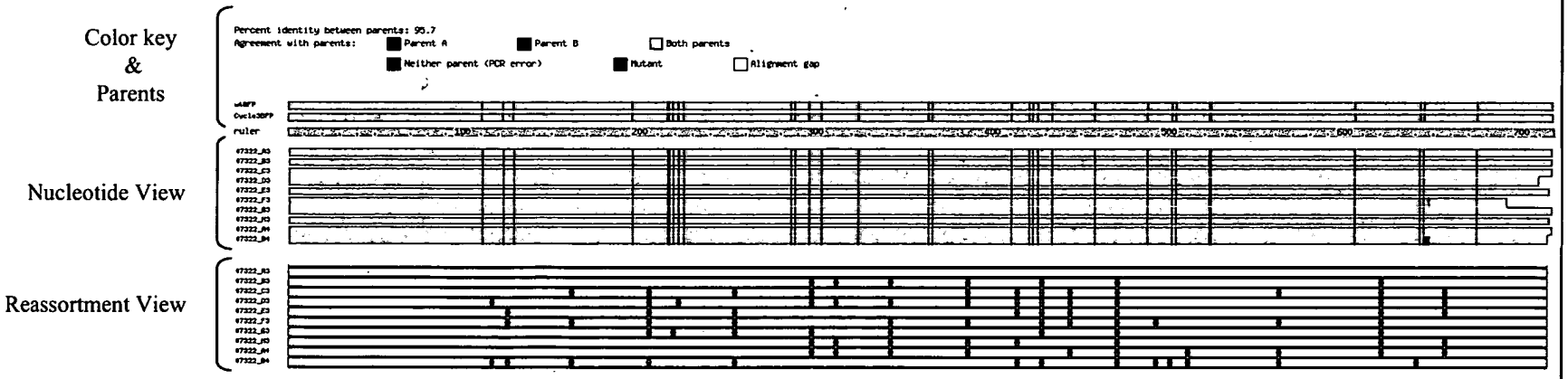
No GRAMMR Treatment - Negative Control



(CEL I omitted from GRAMMR reaction)

Zero Crossovers in 10 Progeny

GRAMMR Treatment



92 Crossovers (>9/gene; >12/kilobase)